(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
4 January 2001 (04.01.2001)

PCT

(10) International Publication Number
WO 01/00878 A2

(51) International Patent Classification[7]: C12Q 1/68, G01N 33/48, C12N 15/18, 15/63, 1/21, 5/00, C07K 14/475, A61K 38/18

(21) International Application Number: PCT/US00/18085

(22) International Filing Date: 29 June 2000 (29.06.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/343,671 30 June 1999 (30.06.1999) US

(71) Applicant: MILLENNIUM PHARMACEUTICALS, INC. [US/US]; 75 Sidney Street, Cambridge, MA 02139 (US).

(72) Inventor: GEARING, David, P.; 23 Standish Road, Wellesley, MA 02146 (US).

(74) Agents: MANDRAGOURAS, Amy, E.; Lahive & Cockfield, LLP, 28 State Street, Boston, MA 02109 et al. (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/00878 A2

(54) Title: A NOVEL VASCULAR ENDOTHELIAL GROWTH FACTOR FAMILY MEMBER AND USES THEREOF

(57) Abstract: The invention provides isolated nucleic acids molecules, designate VEGF-G nucleic acid molecules, which encode novel VEGF family members. The invention also provides antisense nucleic acid molecules, recombinant expression vectors containing VEGF-G nucleic acid molecules, host cells into which the expression vectors have been introduced, and nonhuman transgenic animals in which a VEGF-G gene has been introduced or disrupted. The invention still further provides isolated VEGF-G proteins, fusion proteins, antigenic peptides and anti-VEGF-G antibodies. Diagnostic methods utilizing compositions of the invention are also provided.

A NOVEL VASCULAR ENDOTHELIAL GROWTH FACTOR FAMILY MEMBER AND USES THEREOF

Background of the Invention

The development of a vascular blood supply is required for organogenesis during embryonic development and for reproductive functions in the adult. Angiogenesis is also implicated in physiological and pathophysiological processes such as wound healing and tumor growth and metastasis. Angiogenic factors and growth factors regulate a variety of cellular functions including proliferation, differentiation, migration, and morphogenesis during development.

Vascular endothelial growth factor (VEGF/VEGF-A) is a potent endothelial cell mitogen, survival factor, chemotactic factor and angiogenic factor. The VEGF protein family comprises the structurally and functionally related proteins VEGF-A, VEGF-B/VEGF related factor (VEGF-B/VRF), VEGF-C/VEGF related protein (VEGF-C/VRP), c-fos induced growth factor (FIGF), and placenta growth factor (PlGF) (Erikson, U. *et al.* (1999) *Curr. Topics Microbiol. Immunol.*, 237:4).

VEGF-A is a secreted, dimeric glycoprotein, distantly related in structure to the platelet derived growth factors (PDGF), PDGF-A and PDGF-B. An amino acid sequence alignment of VEGF family members indicates the presence of a core region containing conserved cysteine residues important for inter- and intramolecular disulphide bonding (Erikson, U. *et al.* (1999) *Curr. Topics Microbiol. Immunol.*, 237:41). Several VEGF proteins are also heparin binding growth factors, thus, their activity may be regulated by interactions with cell surface and/or extracellular matrix-associated heparan sulfate proteoglycans (Erikson, U. *et al.* (1999) *Curr. Topics Microbiol. Immunol.*, 237:41; Neufeld, G. *et al.*, (1999) *FASEB J.*, 13:9).

Cell, *e.g.*, endothelial cell, responses are mediated, at least in part, by the binding of VEGF ligands to cell surface VEGF receptors. VEGF receptors include the high affinity tyrosine kinase receptors, Flt and KDR/flk (Neufeld, G. *et al.*, (1999) *FASEB J.*, 13:9; Zachary, I. (1998) *Int. J. Biochem. Cell Biol.*, 30:1169). Additionally, the VEGF-A_{165} protein isoform appears to interact with neuropilin-1, a non-protein tyrosine kinase receptor for the semaphorin family (Neufeld, G. *et al.*, (1999) *FASEB J.*, 13:9; Zachary,

I. (1998) *Int. J. Biochem. Cell Biol.*, 30:1169). The binding of VEGF proteins to their cognate receptors initiates intracellular signalling pathways which result in the activation of MAP kinases and protein kinase C (PKC), phosphorylation of proteins such as focal adhesion kinase and paxillin, and generation of second messengers such as inositol 1,4,5-trisphosphate (IP3) and calcium (Zachary, I. (1998) *Int. J. Biochem. Cell Biol.*, 30:1169).

Summary of the Invention

The present invention is based, at least in part, on the discovery of novel vascular endothelial growth factor family members, referred to herein as "Vascular Endothelial Growth Factor-G" or "VEGF-G" nucleic acid and protein molecules. The VEGF-G molecules of the present invention are useful as modulating agents, or as targets for developing modulating agents to regulate a variety of cellular (*e.g.*, endothelial cell) processes, *e.g.*, cell proliferation, differentiation, migration, angiogenesis, and wound repair. Accordingly, in one aspect, this invention provides isolated nucleic acid molecules encoding VEGF-G proteins or biologically active portions thereof, as well as nucleic acid fragments suitable as primers or hybridization probes for the detection of VEGF-G-encoding nucleic acids.

In one embodiment, a VEGF-G nucleic acid molecule of the invention is at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 86%, 90%, 95%, 98%, or more identical to the nucleotide sequence (*e.g.*, to the entire length of the nucleotide sequence) shown in SEQ ID NO:1, 3, 10 or 12, or a complement thereof.

In another embodiment, the isolated nucleic acid molecule includes the nucleotide sequence shown SEQ ID NO:1 or 3, or a complement thereof. In another embodiment, the nucleic acid molecule includes SEQ ID NO:3 and nucleotides 1-212 of SEQ ID NO:1. In another embodiment, the nucleic acid molecule includes SEQ ID NO:3 and nucleotides 1326-3853 of SEQ ID NO:1. In another embodiment, the nucleic acid molecule consists of the nucleotide sequence shown in SEQ ID NO:1 or 3. In another embodiment, the nucleic acid molecule includes a fragment of at least 549 nucleotides (*e.g.*, 549 contiguous nucleotides) of the nucleotide sequence of SEQ ID NO:1or 3, or a complement thereof.

In another embodiment, the isolated nucleic acid molecule includes the nucleotide sequence shown SEQ ID NO:10 or 12, or a complement thereof. In another embodiment, the nucleic acid molecule includes SEQ ID NO:12 and nucleotides 1-164 of SEQ ID NO:10. In another embodiment, the nucleic acid molecule includes SEQ ID NO:12 and nucleotides 1278-3121 of SEQ ID NO:10. In another embodiment, the nucleic acid molecule consists of the nucleotide sequence shown in SEQ ID NO:10 or 12. In another embodiment, the nucleic acid molecule includes a fragment of at least 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000 or more nucleotides (*e.g.*, contiguous nucleotides) of the nucleotide sequence of SEQ ID NO:10 or 12, or a complement thereof. In still another embodiment, a VEGF-G nucleic acid molecule includes a nucleotide sequence encoding a protein having an amino acid sequence sufficiently homologous to the amino acid sequence of SEQ ID NO:2 or 11. In one embodiment, a VEGF-G nucleic acid molecule includes a nucleotide sequence encoding a protein having an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or more identical to the entire length of the amino acid sequence of SEQ ID NO:2 or 11.

In another embodiment, an isolated nucleic acid molecule encodes the amino acid sequence of human or mouse VEGF-G. In yet another embodiment, the nucleic acid molecule includes a nucleotide sequence encoding a protein having the amino acid sequence of SEQ ID NO:2 or 11.

In yet another embodiment, the nucleic acid molecule is at least 50, 100, 200, 300, 400, 500, 549, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2273, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000 nucleotides in length. In a further preferred embodiment, the nucleic acid molecule is at least 50, 100, 200, 300, 400, 500, 549, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2273, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000 nucleotides in length and encodes a protein having a VEGF-G activity as described herein.

Another embodiment of the invention features nucleic acid molecules, preferably VEGF-G nucleic acid molecules, which specifically detect VEGF-G nucleic acid molecules relative to nucleic acid molecules encoding non-VEGF-G proteins. For example, in one embodiment, such a nucleic acid molecule is at least 50, 60, 70, 80, 90, 100, 150, 200, 300, 400, 500, 549, 549-600, 600-650, 650-700, 700-750, 750-800, 800-850, 850-900, 900-950, 950-1000, 1000-1100, 1100-1200, 1200-1300, 1300-1400, 1400-1500, 1500-1600, 1600-1700, 1700-1800, 1800-1900, 1900-2000, 2000-2200, 2200-2272, 2273, 2273-2400, 2400-2600, 2600-2800, 2800-3000 or more nucleotides in length and hybridizes under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence shown in SEQ ID NO:1 or 10.

In certain embodiments, the nucleic acid molecules are at least 15 (*e.g.*, contiguous) nucleotides in length and hybridize under stringent conditions to nucleotides 1-17 and 310-3283 of SEQ ID NO:1. In another embodiment, the nucleic acid molecules comprise nucleotides 1-17 and 310-3283 of SEQ ID NO:1. In yet another embodiment, the nucleic acid molecules consist of nucleotides 1-17 and 310-3283 of SEQ ID NO:1. In another embodiment, the nucleic acid molecules are at least 15 (*e.g.*, contiguous) nucleotides in length and hybridize under stringent conditions to SEQ ID NO:1or 10.

In other embodiments, the nucleic acid molecule encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2 or 11, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule comprising SEQ ID NO:1, 3, 10 or 12 under stringent conditions.

Another embodiment of the invention provides an isolated nucleic acid molecule which is antisense to a VEGF-G nucleic acid molecule, *e.g.*, the coding strand of a VEGF-G nucleic acid molecule.

In a related aspect, the invention provides a vector comprising a VEGF-G nucleic acid molecule. In certain embodiments, the vector is a recombinant expression vector. In another embodiment, the invention provides a host cell containing a vector of the invention. In yet another embodiment, the invention provides a host cell containing a nucleic acid molecule of the invention. The invention also provides a method for producing a protein, preferably a VEGF-G protein, by culturing in a suitable medium, a

host cell, *e.g.*, a mammalian host cell, such as a non-human mammalian cell, of the invention containing a recombinant expression vector, such that the protein is produced.

Another aspect of this invention features isolated or recombinant VEGF-G proteins and polypeptides. In one embodiment, the isolated VEGF-G protein includes at least one VEGF/PDGF superfamily variant motif. In one embodiment, the isolated VEGF-G protein includes at least one VEGF-G disulfide knot-like domain. In another embodiment, the isolated VEGF-G proteins includes at least one CUB domain. In a further embodiment, the isolated VEGF-G protein includes a signal sequence. In yet another embodiment, the isolated VEGF-G protein includes at least one VEGF/PDGF superfamily variant motif and at least one CUB domain. In yet another embodiment, the isolated VEGF-G protein includes at least one VEGF-G disulfide knot-like domain and at least one CUB domain. In another embodiment, the isolated VEGF-G protein includes at least one VEGF/PDGF superfamily variant motif and a signal sequence. In another embodiment, the isolated VEGF-G protein includes at least one VEGF-G disulfide knot-like domain and a signal sequence. In a further embodiment, the isolated VEGF-G protein includes at least one CUB domain and a signal sequence. In yet another embodiment, the isolated VEGF-G protein includes at least one VEGF/PDGF superfamily variant motif, at least one CUB domain, and a signal sequence. In yet another embodiment, the isolated VEGF-G protein includes at least one VEGF-G disulfide knot-like domain, at least one CUB domain, and a signal sequence.

In other embodiments, the VEGF-G protein of the invention has an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or more identical to the amino acid sequence of SEQ ID NO:2 or 11. In another embodiment, the VEGF-G protein includes a VEGF/PDGF superfamily variant motif and a CUB domain, and has an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or more identical to the amino acid sequence of SEQ ID NO:2 or 11. In another embodiment, the VEGF-G protein includes a VEGF-G disulfide knot-like domain and a CUB domain, and has an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or more identical to the amino acid sequence of SEQ ID NO:2 or 11.

G nucleic acid molecule, protein or polypeptide such that the presence of a VEGF-G nucleic acid molecule, protein or polypeptide is detected in the biological sample.

In another aspect, the present invention provides a method for detecting the presence of VEGF-G activity in a biological sample by contacting the biological sample with an agent capable of detecting an indicator of VEGF-G activity such that the presence of VEGF-G activity is detected in the biological sample.

In another aspect, the invention provides a method for modulating VEGF-G activity comprising contacting a cell capable of expressing VEGF-G with an agent that modulates VEGF-G activity such that VEGF-G activity in the cell is modulated. In one embodiment, the agent inhibits VEGF-G activity. In another embodiment, the agent stimulates VEGF-G activity. In one embodiment, the agent is an antibody that specifically binds to a VEGF-G protein. In another embodiment, the agent modulates expression of VEGF-G by modulating transcription of a VEGF-G gene or translation of a VEGF-G mRNA. In yet another embodiment, the agent is a nucleic acid molecule having a nucleotide sequence that is antisense to the coding strand of a VEGF-G mRNA or a VEGF-G gene.

Another aspect of the present invention features methods to treat a subject having a disorder characterized by aberrant VEGF-G protein or nucleic acid expression or activity by administering an agent which is a VEGF-G modulator to the subject. In one embodiment, the VEGF-G modulator is a VEGF-G protein. In another embodiment the VEGF-G modulator is a VEGF-G nucleic acid molecule. In yet another embodiment, the VEGF-G modulator is a peptide, peptidomimetic, or other small molecule. In one embodiment, the disorder characterized by aberrant VEGF-G protein or nucleic acid expression is a disorder associated with deregulated cell growth such as a proliferative or differentiative disorder, including cancer, *e.g.*, carcinoma, sarcoma, or leukemia, and hypertrophic bone disorders, *e.g.*, opsismodysplasia; and/or a disorder involving aberrant angiogenesis and/or vascularity, *e.g.*, tumor angiogenesis and metastasis, diabetic retinopathy, macular degeneration, psoriasis, endometriosis, Grave's disease, ischemic disease (*e.g.*, atherosclerosis), and chronic inflammatory disease (*e.g.*, rheumatoid arthritis).

In another embodiment, the VEGF-G proteins of the invention play a role in endothelial cell growth, *e.g.*, the regulation of cell proliferation, differentiation, migration, and apoptosis.

In other embodiments, the VEGF-G proteins of the invention are encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, 3, 10 or 12.

In a further embodiment, the invention features an isolated VEGF-G protein which is encoded by a nucleic acid molecule consisting of a nucleotide sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 86%, 90%, 95%, 98% or more identical to a nucleotide sequence of SEQ ID NO:1, 3, 10 or 12, or a complement thereof. This invention further features an isolated VEGF-G protein which is encoded by a nucleic acid molecule consisting of a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, 3, 10 or 12, or a complement thereof. In still another embodiment, the VEGF-G protein has the amino acid sequence of SEQ ID NO:2 or 11.

In another embodiment, the invention features fragments of the protein having the amino acid sequence of SEQ ID NO:2 or 11, wherein the fragment comprises at least 15 amino acids (*e.g.*, contiguous amino acids) of the amino acid sequence of SEQ ID NO:2 or 11.

The proteins of the present invention or portions thereof, *e.g.*, biologically active portions thereof, can be operatively linked to a non-VEGF-G polypeptide (*e.g.*, heterologous amino acid sequences) to form fusion proteins. In addition, the VEGF-G proteins or biologically active portions thereof can be incorporated into pharmaceutical compositions, which optionally include pharmaceutically acceptable carriers.

The invention further features antibodies, such as monoclonal or polyclonal antibodies, that specifically bind proteins of the invention, preferably VEGF-G proteins.

In another aspect, the present invention provides a method for detecting the presence of a VEGF-G nucleic acid molecule, protein or polypeptide in a biological sample by contacting the biological sample with an agent capable of detecting a VEGF-

The present invention also provides a diagnostic assay for identifying the presence or absence of a genetic alteration characterized by at least one of (i) aberrant modification or mutation of a gene encoding a VEGF-G protein; (ii) mis-regulation of the gene; and (iii) aberrant post-translational modification of a VEGF-G protein, wherein a wild-type form of the gene encodes a protein with a VEGF-G activity.

In another aspect the invention provides a method for identifying a compound that binds to or modulates the activity of a VEGF-G protein, by providing an indicator composition comprising a VEGF-G protein having VEGF-G activity, contacting the indicator composition with a test compound, and determining the effect of the test compound on VEGF-G activity in the indicator composition to identify a compound that modulates the activity of a VEGF-G protein.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

Brief Description of the Drawings

Figure 1 depicts the cDNA sequence and predicted amino acid sequence of human VEGF-G. The nucleotide sequence corresponds to nucleic acids 1 to 3853 of SEQ ID NO:1. The amino acid sequence corresponds to amino acids 1 to 370 of SEQ ID NO: 2. The coding region without the 5' and 3' untranslated regions of the human VEGF-G gene is shown in SEQ ID NO:3.

Figure 2 depicts the results of a search which was performed against the HMM database and which resulted in the identification of a "CUB domain" in the human VEGF-G protein.

Figure 3 depicts a local alignment of the human VEGF-G protein with the human VEGF-C protein using the GAP program in the GCG software package, using a Blossum 62 matrix and a gap weight of 5 and a length weight of 2. The results showed a 30.357% identity between the two sequences.

Figure 4 depicts a local alignment of the human VEGF-G protein with the human VEGF-A protein using the GAP program in the GCG software package, using a Blossum 62 matrix and a gap weight of 5 and a length weight of 2. The results showed a 26.042% identity between the two sequences.

Figure 5 depicts a local alignment of the human VEGF-G protein with the human TANGO 128 protein using the GAP program in the GCG software package, using a Blossum 62 matrix and a gap weight of 12 and a length weight of 4. The results showed a 51.203% identity between the two sequences.

Figure 6 depicts a global alignment of the human VEGF-G protein with the human TANGO 128 protein using the ALIGN program (version 2.0), using a PAM120 scoring matrix, and a gap length penalty of 12 and a gap penalty of 4. The results showed a 42.4% identity between the two sequences.

Figure 7 depicts a global alignment of the human VEGF-G nucleic acid sequence with the human TANGO 128 nucleic acid sequence using the ALIGN program (version 2.0), using a PAM120 scoring matrix, and a gap length penalty of 12 and a gap penalty of 4. The results showed a 51.9% identity between the two sequences.

Figure 8 depicts the cDNA sequence and predicted amino acid sequence of mouse VEGF-G. The nucleotide sequence corresponds to nucleic acids 1 to 3121 of SEQ ID NO:10. The amino acid sequence corresponds to amino acids 1 to 370 of SEQ ID NO: 11. The coding region without the 5' and 3' untranslated regions of the mouse VEGF-G gene is shown in SEQ ID NO:12.

Figure 9, panel A depicts the results of a search which was performed against the HMM database and which resulted in the identification of a "CUB domain" in the mouse VEGF-G protein; panel B depicts the results of a search which was performed against the SMART database and which resulted in the identification of a "CUB domain" and a "PDGF domain" in the mouse VEFG-G protein.

Figure 10 depicts a global alignment of the open reading frame of the human VEGF-G gene with the open reading frame of the mouse VEGF-G gene using the GAP program in the GCG software package, using a nwsgapdna.cmp matrix and a gap weight of 12 and a length weight of 4. The results showed a 85.586% identity between the two sequences.

Figure 11 depicts a global alignment of the human VEGF-G protein with the mouse VEGF-G protein using the GAP program in the GCG software package, using a Blosum 62 matrix and a gap weight of 12 and a length weight of 4. The results showed a 84.865% identity between the two sequences.

Detailed Description of the Invention

The present invention is based, at least in part, on the discovery of novel VEGF family members, referred to herein as "Vascular Endothelial Growth Factor-G" or "VEGF-G" nucleic acid and protein molecules. VEGF proteins modulate the proliferation, motility, differentiation and survival of cells, such as endothelial cells or monocytes. VEGF molecules regulate a broad range of cell, *e.g.*, endothelial cell, processes including mitogenesis, permeability, vascular tone, and the synthesis of vasoactive molecules (Zachary, I. (1998) *Int. J. Biochem. Cell Biol.*, 30:1169). VEGF proteins are important in vasculogenesis and angiogenesis, both in normal physiology and the pathophysiology of disease states.

The VEGF-G molecules of the present invention are predicted to be growth regulatory proteins that function to modulate cell proliferation, differentiation, motility, and apoptosis. Thus, the VEGF-G molecules of the present invention may play a role in cellular growth signalling mechanisms. As used herein, the term "cellular growth signalling mechanism" includes signal transmissions from cell receptors, *e.g.*, growth factor receptors, which regulate one or more of the following: 1) cell transversal through the cell cycle, 2) cell differentiation, 3) cell migration and patterning, and 4) programmed cell death. Throughout development and in the adult organism, cell fate and activity is determined, in part, by extracellular and intracellular stimuli, *e.g.*, growth factors, angiogenic factors, chemotactic factors, neurotrophic factors, cytokines, and hormones. These stimuli act on their target cells by initiating signal transduction cascades that alter the pattern of gene expression and metabolic activity so as to mediate the appropriate cellular response. The VEGF-G molecules of the present invention are predicted to be involved in the initiation or modulation of cellular signal transduction pathways that modulate endothelial cell growth, differentiation, migration and/or apoptosis. Thus, the VEGF-G molecules, by participating in cellular growth signalling

mechanisms, may modulate cell behavior and act as therapeutic agents for controlling cellular proliferation, differentiation, migration, and apoptosis.

Altered expression of factors (*e.g.*, a VEGF-G molecule) involved in the regulation of signalling pathways associated with cell (*e.g.*, endothelial cell) growth, differentiation, migration, and apoptosis can lead to perturbed cellular proliferation, which in turn can lead to cellular proliferative and/or differentiative disorders. As used herein, a "cellular proliferative disorder" includes a disorder, disease, or condition characterized by a deregulated, *e.g.*, upregulated or downregulated, growth response. As used herein, a "cellular differentiative disorder" includes a disorder, disease, or condition characterized by aberrant cellular differentiation. Thus, the VEGF-G molecules can act as novel diagnostic targets and therapeutic agents for controlling cellular proliferative and/or differentiative disorders. Examples of cellular proliferative and/or differentiative disorders include cancer, *e.g.*, carcinoma, sarcoma, or leukemia; and disorders involving aberrant angiogenesis and/or vascularity, *e.g.*, tumor angiogenesis and metastasis, diabetic retinopathy, macular degeneration, psoriasis, endometriosis, Grave's disease, ischemic disease (*e.g.*, atherosclerosis), and chronic inflammatory diseases (*e.g.*, rheumatoid arthritis).

VEGF-G-associated or related disorders also include disorders affecting tissues in which VEGF-G protein is expressed.

The term "family" when referring to the protein and nucleic acid molecules of the invention is intended to mean two or more proteins or nucleic acid molecules having a common structural domain or motif and having sufficient amino acid or nucleotide sequence homology as defined herein. Such family members can be naturally or non-naturally occurring and can be from either the same or different species. For example, a family can contain a first protein of human origin as well as other distinct proteins of human origin, or alternatively, can contain homologues of non-human origin, *e.g.*, rat or mouse proteins. Members of a family can also have common functional characteristics.

For example, members of the VEGF-G family of proteins include at least one "VEGF/PDGF superfamily variant motif" in the protein molecule or the nucleic acid molecule encoding the protein molecule. As used herein, the term " VEGF/PDGF superfamily variant motif" includes a protein domain having an amino acid sequence of

about 60-130 amino acid residues. Preferably, a VEGF/PDGF superfamily variant motif includes at least about 70-120, more preferably about 80-110 amino acid residues, and most preferably about 90-100 amino acid residues. The VEGF/PDGF superfamily variant motif is characterized by conserved cysteine residues that form intra- and inter-chain disulfide bonds which may affect the structural integrity of the protein, and in one embodiment has the following signature pattern:

[RK]-X(2-3)-C-X-[PA]-X(19-20)-P-X-C-X(4)-R-C-G-G-N-C-[GA]-C-X(6-7)-C-X(30-45)-C-X-C (SEQ ID NO:4)

The signature patterns or consensus patterns described herein are described according to the following designation: all amino acids are indicated according to their universal single letter designation; "x" designates any amino acid; x(n) designates n number of amino acids, *e.g.*, x (2) designates any two amino acids, *e.g.*, x (1-3) designates any of one to three amino acids; and, amino acids in brackets indicates any one of the amino acids within the brackets, *e.g.*, [RK] indicates any of one of either R (arginine) or K (lysine). VEGF-G has such a signature pattern at about amino acids 268 to 362 of SEQ ID NO:2, and at about amino acids 268-362 of SEQ ID NO:11. Another VEGF-G family member, human TANGO 128 (U.S. Patent Application No. 09/223,546, filed Dec. 30, 1998), has a VEGF/PDGF superfamily variant motif, suggesting that VEGF-G and TANGO 128 may represent a novel variant subclass of the VEGF/PDGF family of molecules.

In another embodiment, the VEGF/PDGF superfamily variant motif has the following signature pattern:

[RK]-X(2-3)-C-X-[PA]-X(19-20)-P-X-C-X(4)-R-C-G-G-N-C-[GA]-C-X(6-7)-C-X(32-43)-C-X-C (SEQ ID NO:5)

VEGF-G has such a signature pattern at about amino acids 268 to 362 of SEQ ID NO:2, and at about amino acids 268 to 362 of SEQ ID NO:11.

In yet another embodiment, the VEGF/PDGF superfamily variant motif has the following signature pattern:

[RK]-X(2-3)-C-X-[PA]-X(19-20)-P-X-C-X(4)-R-C-G-G-N-C-[GA]-C-X(7)-C-X-C-X(38-41)-C-X-C-X-C (SEQ ID NO:6)

VEGF-G has such a signature pattern at about amino acids 268 to 364 of SEQ ID NO:2, and at about amino acids 268 to 364 of SEQ ID NO:11.

In still another embodiment, a member of the VEGF-G family of molecules includes a VEGF-G disulfide knot-like domain. A VEGF-G disulfide knot-like domain has the following signature pattern :

C-X(3)-C-[GA]-C (SEQ ID NO:7)

VEGF-G has such a signature pattern at about amino acids 302 to 308 of SEQ ID NO:2, and at about amino acids 302-308 of SEQ ID NO:11.

In a preferred embodiment, the VEGF-G disulfide knot-like domain has the following signature pattern :

C-G-G-N-C-[GA]-C (SEQ ID NO:8)

VEGF-G has such a signature pattern at about amino acids 302 to 308 of SEQ ID NO:2, and at about amino acids 302-308 of SEQ ID NO:11.

In another preferred embodiment, a member of this novel subfamily of VEGF proteins has a VEGF/PDGF superfamily variant motif which includes at least about 60-130 amino acid residues and has at least about 50-60% identity with a VEGF/PDGF superfamily variant motif of VEGF-G (*e.g.*, residues 268-362 of SEQ ID NO:2 or 11). Preferably, a VEGF/PDGF superfamily variant motif includes at least about 70-120 amino acid residues, or about 80-110 amino acid residues, or 90-100 amino acid residues, and has at least 60-70% identity, preferably about 70-80%, or more preferably about 80-90%, identity with a VEGF/PDGF superfamily variant motif of VEGF-G (*e.g.*, residues 268-362 of SEQ ID NO:2 or 11).

Accordingly, VEGF-G proteins having at least 50-60% identity, preferably about 60-70%, more preferably about 70-80%, or about 80-90% identity with a VEGF/PDGF superfamily variant motif of human or mouse VEGF-G are within the scope of the invention.

VEGF-G family members can be identified based on the presence of a "CUB domain" in the protein or the nucleic acid molecule encoding the protein. As used herein, the term "CUB domain" includes a protein domain having an amino acid sequence of about 90-140 amino acid residues and having a bit score for the alignment of the sequence to the CUB domain (HMM) of at least about 80. Preferably, a CUB domain includes at least about 100-130, more preferably about 110-120 amino acid residues, or 113-117 amino acid residues, and has a bit score for the alignment of the

sequence to the CUB domain (HMM) of at least about 85, 90, 100, 120 or greater. The CUB domain (HMM) has been assigned the PFAM Accession PF00431 (http://genome.wustl.edu/Pfam/.html). The CUB domain is an extracellular domain associated with various developmentally regulated proteins and as such is likely to be involved in developmental processes. A CUB domain contains conserved cysteine residues which are likely to form disulfide bonds that affect protein structure.

To identify the presence of a CUB domain in a VEGF-G protein, and make the determination that a protein of interest has a particular profile, the amino acid sequence of the protein is searched against a database of HMMs (*e.g.*, the Pfam database, release 2.1) using the default parameters (http://www.sanger.ac.uk/Software/Pfam/HMM_search). For example, the hmmsf program, which is available as part of the HMMER package of search programs, is a family specific default program for MILPAT0063 and a score of 15 is the default threshold score for determining a hit. Alternatively, the threshold score for determining a hit can be lowered (*e.g.*, to 8 bits). A description of the Pfam database can be found in Sonhammer *et al.* (1997) *Proteins* 28(3):405-420 and a detailed description of HMMs can be found, for example, in Gribskov *et al.*(1990) *Meth. Enzymol.* 183:146-159; Gribskov *et al.*(1987) *Proc. Natl. Acad. Sci. USA* 84:4355-4358; Krogh *et al.*(1994) *J. Mol. Biol.* 235:1501-1531; and Stultz *et al.*(1993) *Protein Sci.* 2:305-314, the contents of which are incorporated herein by reference. A search was performed against the HMM database resulting in the identification of a CUB domain in the amino acid sequence of human VEGF-G at about residues 53-167 of SEQ ID NO:2 (see Figure 2). A CUB domain was also identified in the amino acid sequence of mouse VEGF-G at about residues 53-167 of SEQ ID NO:11 (see Figure 9).In one embodiment, a member of the VEGF-G family of molecules includes a CUB domain with the following signature pattern:

G-X(3)-S-P-X(2)-P-X(2)-Y-P-X(6)-W-X(27-28)-C-X-Y-D-X(13-15)-G-X(2)-C-G-X(3-4)-P-X(16)-D-X(6-7)-G-F (SEQ ID NO:9)

VEGF-G has such a signature pattern at about amino acids 62 to 163 of SEQ ID NO:2, and at about amino acids 62 to 163 of SEQ ID NO:11.

In a further preferred embodiment, a member of this novel subfamily of VEGF proteins has a CUB domain includes at least about 70-130 amino acid residues and has at least about 50-60% identity with a CUB domain of VEGF-G (*e.g.*, residues 62-163 of SEQ ID NO:2 or 11). Preferably, a CUB domain includes at least about 80-120 amino acid residues, or about 90-110 amino acid residues, or 98-104 amino acid residues, and has at least 60-70% identity, preferably about 70-80%, more preferably about 80-90% identity with a CUB domain of VEGF-G (*e.g.*, residues 62-163 of SEQ ID NO:2 or 11).

Accordingly, VEGF-G proteins having at least 50-60% identity, preferably about 60-70%, more preferably about 70-80%, or about 80-90% identity with a CUB domain of human or mouse VEGF-G are within the scope of the invention.

VEGF-G family members can be identified based on the presence of a signal sequence. As used herein, a "signal sequence" includes a peptide containing about 19 amino acids which occurs at the N-terminus of secretory and membrane bound proteins and which contains a large number of hydrophobic amino acid residues. For example, a signal sequence contains at least about 10-30 amino acid residues, preferably about 15-25 amino acid residues, more preferably about 18-20 amino acid residues, and more preferably about 19 amino acid residues, and has at least about 35-65%, preferably about 38-50%, and more preferably about 40-45% hydrophobic amino acid residues (*e.g.*, Valine, Leucine, Isoleucine or Phenylalanine). Such a "signal sequence", also referred to in the art as a "signal peptide", serves to direct a protein containing such a sequence to a lipid bilayer, and is cleaved in secreted and membrane bound proteins. A signal sequence was identified in the amino acid sequence of human VEGF-G at about amino acids 1-19 of SEQ ID NO:2. A signal sequence was also identified in the amino acid sequence of mouse VEGF-G at about amino acids 1-23 of SEQ ID NO:11.

In yet another embodiment, VEGF-G family members include at least one or more N-glycosylation sites. Predicted N-glycosylation sites are found, for example, from about amino acids 276-279 of SEQ ID NO:2, and from about amino acid residues 14-17 and 276-279 of SEQ ID NO:11.

In another embodiment, VEGF-G family members include at least one N-myristoylation site. Predicted N-myristoylation sites are found, for example, from about amino acid residues 100-105, 192-197, and 303-308 of SEQ ID NO:2, and from about amino acid residues 100-105 and 303-308 of SEQ ID NO:11.

In another embodiment, VEGF-G family members include at least one cAMP and cGMP dependent protein kinase phosphorylation site. Predicted cAMPand cGMP dependent protein kinase phosphorylation sites are found, for example, from about amino acid residues 268-271 of SEQ ID NO:2, and from about amino acid residues 268-271 of SEQ ID NO:11.

In another embodiment, VEGF-G family members include at least one protein kinase C phosphorylation site. Predicted protein kinase C phosphorylation sites are found, for example, from about amino acid residues 17-19, 29-31, 66-68, 80-82, 150-152, 243-245, 273-275, 320-322, 323-325, and 365-367 of SEQ ID NO:2, and from about amino acid residues 29-31, 66-68, 141-143, 150-152, 273-275, 320-322, 323-325, and 365-367 of SEQ ID NO:11.

In another embodiment, VEGF-G family members include at least one casein kinase II phosphorylation site. Predicted casein kinase II phosphorylation sites are found, for example, from about amino acid residues 17-20, 168-171, 181-184, 199-202, 219-222, 231-234, 250-253, and 256-259 of SEQ ID NO:2, and from about amino acid residues 168-171, 181-184, 199-202, 213-216, 219-222, 231-234, 250-253, and 256-259 of SEQ ID NO:11.

In another embodiment, VEGF-G family members include at least one tyrosine kinase phosphorylation site. Predicted tyrosine kinase phosphorylation sites are found, for example, from about amino acid residues 262-270 of SEQ ID NO:2, and from about amino acid residues 262-270 of SEQ ID NO:11.

Isolated proteins of the present invention, preferably VEGF-G proteins, have an amino acid sequence sufficiently homologous to the amino acid sequence of SEQ ID NO:2 or 11, or are encoded by a nucleotide sequence sufficiently homologous to SEQ ID NO:1, 3, 10 or 12. As used herein, the term "sufficiently homologous" refers to a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent (*e.g.*, an amino acid residue which has a similar side chain)

amino acid residues or nucleotides to a second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide sequences share common structural domains or motifs and/or a common functional activity. For example, amino acid or nucleotide sequences which share common structural domains have at least 50% homology, preferably 60% homology, more preferably 70%-80%, and even more preferably 90-95% homology across the amino acid sequences of the domains and contain at least one and preferably two structural domains or motifs, are defined herein as sufficiently homologous. Furthermore, amino acid or nucleotide sequences which share at least 50%, preferably 60%, more preferably 70-80%, or 90-95% homology and share a common functional activity are defined herein as sufficiently homologous.

As used interchangeably herein, an "VEGF-G activity", "biological activity of VEGF-G" or "functional activity of VEGF-G", refers to an activity exerted by a VEGF-G protein, polypeptide or nucleic acid molecule on a VEGF-G responsive cell or on a VEGF-G protein substrate, as determined *in vivo* or *in vitro*, according to standard techniques. In one embodiment, a VEGF-G activity is a direct activity, such as an association with a VEGF-G target molecule. As used herein, a "target molecule" or "binding partner" is a molecule with which a VEGF-G protein binds or interacts in nature, such that VEGF-G-mediated function is achieved. A VEGF-G target molecule can be a non-VEGF-G molecule or a VEGF-G protein or polypeptide of the present invention. In an exemplary embodiment, a VEGF-G target molecule is a VEGF-G substrate or receptor, *e.g.*, flt or KDR/flk. A VEGF-G activity can also be an indirect activity, such as a cellular signalling activity mediated by interaction of the VEGF-G protein with a VEGF-G substrate or receptor, *e.g.*, flt or KDR/flk. Preferably, a VEGF-G activity is the ability to act as a growth regulatory factor and to modulate cell proliferation, differentiation, migration, apoptosis, and/or angiogenesis.

Accordingly, another embodiment of the invention features isolated VEGF-G proteins and polypeptides having a VEGF-G activity. Preferred proteins are VEGF-G proteins including at least one VEGF/PDGF superfamily variant motif, and, preferably, having a VEGF-G activity. Further preferred proteins include at least one VEGF/PDGF superfamily variant motif, and are, preferably, encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions

to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, 3, 10 or 12. Preferred proteins are VEGF-G proteins including at least one VEGF-G disulfide knot-like domain, and, preferably, having a VEGF-G activity. Further preferred proteins include at least one VEGF-G disulfide knot-like domain, and are, preferably, encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, 3, 10 or 12. Other preferred proteins are VEGF-G proteins including at least one CUB domain, and, preferably, having a VEGF-G activity. Additional preferred proteins include at least one CUB domain, and are, preferably, encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, 3, 10 or 12. Preferred proteins are VEGF-G proteins including at least one signal sequence, and, preferably, having a VEGF-G activity. Further preferred proteins include at least one signal sequence, and are, preferably, encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, 3, 10 or 12. Additional preferred proteins are VEGF-G proteins including at least one VEGF/PDGF superfamily variant motif and at least one CUB domain, and preferably, having a VEGF-G activity. Further preferred proteins include at least one VEGF/PDGF superfamily variant motif and at least one CUB domain and are, preferably, encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, 3, 10 or 12. Other preferred proteins are VEGF-G proteins including at least one VEGF/PDGF superfamily variant motif, at least one CUB domain, and at least one signal sequence and preferably, having a VEGF-G activity. Further preferred proteins include at least one VEGF/PDGF superfamily variant motif, at least one CUB domain, and at least one signal sequence and are, preferably, encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, 3, 10 or 12. Additional preferred proteins are VEGF-G proteins including at least one VEGF-G disulfide knot-like

domain and at least one CUB domain, and preferably, having a VEGF-G activity. Further preferred proteins include at least one VEGF-G disulfide knot-like domain and at least one CUB domain and are, preferably, encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, 3, 10 or 12. Other preferred proteins are VEGF-G proteins including at least one VEGF-G disulfide knot-like domain, at least one CUB domain, and at least one signal sequence and preferably, having a VEGF-G activity. Further preferred proteins include at least one VEGF-G disulfide knot-like domain, at least one CUB domain, and at least one signal sequence and are, preferably, encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, 3, 10 or 12.

The nucleotide sequence of the isolated human VEGF-G cDNA and the predicted amino acid sequence of the human VEGF-G polypeptide are shown in Figure 1 and in SEQ ID NOs:1 and 2, respectively. The human VEGF-G gene, which is approximately 3853 nucleotides in length, encodes a protein having a molecular weight of approximately 43 kD and which is approximately 370 amino acid residues in length.

The nucleotide sequence of the isolated mouse VEGF-G cDNA and the predicted amino acid sequence of the mouse VEGF-G polypeptide are shown in Figure 8 and in SEQ ID NOs:10 and 11, respectively. The mouse VEGF-G gene, which is approximately 3121 nucleotides in length, encodes a protein having a molecular weight of approximately 43 kD and which is approximately 370 amino acid residues in length. The predicted mature mouse VEGF-G protein lacking the signal sequence has a molecular weight of approximately 40 kD and is approximately 347 amino acid residues in length.

Various aspects of the invention are described in further detail in the following subsections:

I. Isolated Nucleic Acid Molecules

One aspect of the invention pertains to isolated nucleic acid molecules that encode VEGF-G proteins or biologically active portions thereof, as well as nucleic acid

fragments sufficient for use as hybridization probes to identify VEGF-G -encoding nucleic acid molecules (*e.g.*, VEGF-G mRNA) and fragments for use as PCR primers for the amplification or mutation of VEGF-G nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA) and RNA molecules (*e.g.*, mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

The term "isolated nucleic acid molecule" includes nucleic acid molecules which are separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. For example, with regards to genomic DNA, the term "isolated" includes nucleic acid molecules which are separated from the chromosome with which the genomic DNA is naturally associated. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated VEGF-G nucleic acid molecule can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1, 3, 10 or 12, or a portion thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NO:1, 3, 10 or 12 as hybridization probes, VEGF-G nucleic acid molecules can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd. ed., Cold Spring Harbor Laboratory*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Moreover, a nucleic acid molecule encompassing all or a portion of SEQ ID NO:1, 3, 10 or 12 can be isolated by the polymerase chain reaction (PCR) using synthetic oligonucleotide primers designed based upon the sequence of SEQ ID NO:1, 3, 10 or 12.

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to VEGF-G nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

In one embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence shown in SEQ ID NO:1. The sequence of SEQ ID NO:1 corresponds to the human VEGF-G cDNA. This cDNA comprises sequences encoding the human VEGF-G protein (*i.e.*, "the coding region", from nucleotides 213-1325), as well as 5' untranslated sequences (nucleotides 1-212) and 3' untranslated sequences (nucleotides 1326-3853). Alternatively, the nucleic acid molecule can comprise only the coding region of SEQ ID NO:1 (*e.g.*, nucleotides 213-1325, corresponding to SEQ ID NO:3).

In one embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence shown in SEQ ID NO:10. The sequence of SEQ ID NO:10 corresponds to the mouse VEGF-G cDNA. This cDNA comprises sequences encoding the mouse VEGF-G protein (*i.e.*, "the coding region", from nucleotides 165-1277), as well as 5' untranslated sequences (nucleotides 1-164) and 3' untranslated sequences (nucleotides 1278-3121). Alternatively, the nucleic acid molecule can comprise only the coding region of SEQ ID NO:10 (*e.g.*, nucleotides 165-1277, corresponding to SEQ ID NO:12).

In one embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of the nucleotide sequence shown in SEQ ID NO:1, 3, 10 or 12, or a portion of any of these nucleotide sequences. A nucleic acid molecule which is complementary to the nucleotide sequence shown in SEQ ID NO:1, 3, 10 or 12 is one which is sufficiently complementary to the nucleotide sequence

shown in SEQ ID NO:1, 3, 10 or 12, such that it can hybridize to the nucleotide sequence shown in SEQ ID NO:1, 3, 10 or 12, thereby forming a stable duplex.

In one embodiment, an isolated nucleic acid molecule of the present invention comprises a nucleotide sequence which is at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 86%, 90%, 95%, 98% or more homologous to the entire length of the nucleotide sequence shown in SEQ ID NO:1, 3, 10 or 12, or a portion of any of these nucleotide sequences.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the nucleic acid sequence of SEQ ID NO:1, 3, 10 or 12, for example, a fragment which can be used as a probe or primer or a fragment encoding a portion of a VEGF-G protein, *e.g.*, an immunogenic or biologically active portion of a VEGF-G protein. The nucleotide sequence determined from the cloning of the VEGF-G gene allows for the generation of probes and primers designed for use in identifying and/or cloning other VEGF-G family members, as well as VEGF-G homologues from other species. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12 or 15, preferably about 20 or 25, more preferably about 30, 35, 40, 45, 50, 55, 60, 65, or 75 consecutive nucleotides of a sense or antisense sequence of SEQ ID NO:1, 3, 10 or 12, or of a naturally occurring allelic variant or mutant of SEQ ID NO:1, 3, 10 or 12. In an exemplary embodiment, a nucleic acid molecule of the present invention comprises a nucleotide sequence which is greater than 50, 60, 70, 80, 90, 100, 150, 200, 300, 400, 500, 549, 549-600, 600-650, 650-700, 700-750, 750-800, 800-850, 850-900, 900-950, 950-1000, 1000-1100, 1100-1200, 1200-1300, 1300-1400, 1400-1500, 1500-1600, 1600-1700, 1700-1800, 1800-1900, 1900-2000, 2000-2200, 2200-2272, 2273, 2273-2400, 2400-2600, 2600-2800, 2800-3000 or more nucleotides in length and hybridizes under stringent hybridization conditions to a nucleic acid molecule of SEQ ID NO:1, 3, 10 or 12.

Probes based on the VEGF-G nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In certain embodiments, the probe further comprises a label group attached thereto, *e.g.*, the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-

factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissue which misexpress a VEGF-G protein, such as by measuring a level of a VEGF-G-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting VEGF-G mRNA levels or determining whether a genomic VEGF-G gene has been mutated or deleted.

A nucleic acid fragment encoding a "biologically active portion of a VEGF-G protein" can be prepared by isolating a portion of the nucleotide sequence of SEQ ID NO:1, 3, 10 or 12, which encodes a polypeptide having a VEGF-G biological activity (the biological activities of the VEGF-G proteins are described herein), expressing the encoded portion of the VEGF-G protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the VEGF-G protein. A nucleic acid fragment encoding a biologically active portion of a VEGF-G protein, may comprise a nucleotide sequence which is greater than 50, 60, 70, 80, 90, 100, 150, 200, 300, 400, 500, 549, 549-600, 600-650, 650-700, 700-750, 750-800, 800-850, 850-900, 900-950, 950-1000, 1000-1100, 1100-1200, 1200-1300, 1300-1400, 1400-1500, 1500-1600, 1600-1800, 1800-2000, 2000-2200, 2273 or more nucleotides in length.

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence shown in SEQ ID NO:1, 3, 10 or 12 due to degeneracy of the genetic code and thus encode the same VEGF-G proteins as those encoded by the nucleotide sequence shown in SEQ ID NO:1, 3, 10 or 12. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in SEQ ID NO:2 or 11.

In addition to the VEGF-G nucleotide sequences shown in SEQ ID NO:1, 3, 10 or 12, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the VEGF-G proteins may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the VEGF-G genes may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules which include an open reading frame encoding a VEGF-G protein, preferably a mammalian VEGF-G protein, and can further include non-coding regulatory sequences, and introns.

Allelic variants of VEGF-G, *e.g.*, human or mouse VEGF-G, include both functional and non-functional VEGF-G proteins. Functional allelic variants are naturally occurring amino acid sequence variants of the VEGF-G protein within a population that maintain the ability to bind a VEGF-G receptor or substrate, and/or modulate cell growth and migration mechanisms. Functional allelic variants will typically contain only conservative substitution of one or more amino acids of SEQ ID NO:2 or 11, or substitution, deletion or insertion of non-critical residues in non-critical regions of the protein.

Non-functional allelic variants are naturally occurring amino acid sequence variants of the VEGF-G, *e.g.*, human or mouse VEGF-G, protein within a population that do not have the ability to either bind a VEGF-G receptor or substrate, or modulate cell growth or migration mechanisms. Non-functional allelic variants will typically contain a non-conservative substitution, a deletion, or insertion, or premature truncation of the amino acid sequence of SEQ ID NO:2 or 11, or a substitution, insertion, or deletion in critical residues or critical regions of the protein.

The present invention further provides orthologues of the human and mouse VEGF-G proteins. Orthologues of the human and mouse VEGF-G proteins are proteins that are isolated from non-human and non-mouse organisms and possess the same VEGF-G receptor or substrate binding mechanisms, and/or modulation of cell growth or migration mechanisms of the human and mouse VEGF-G proteins. Orthologues of the human and mouse VEGF-G proteins can readily be identified as comprising an amino acid sequence that is substantially homologous to SEQ ID NO:2 or 11.

Moreover, nucleic acid molecules encoding other VEGF-G family members and, thus, which have a nucleotide sequence which differs from the VEGF-G sequences of SEQ ID NO:1, 3, 10 or 12 are intended to be within the scope of the invention. For example, another VEGF-G cDNA can be identified based on the nucleotide sequence of human or mouse VEGF-G. Moreover, nucleic acid molecules encoding VEGF-G proteins from different species, and which, thus, have a nucleotide sequence which differs from the VEGF-G sequences of SEQ ID NO:1, 3, 10 or 12 are intended to be within the scope of the invention. For example, a monkey VEGF-G cDNA can be identified based on the nucleotide sequence of a human or mouse VEGF-G.

Nucleic acid molecules corresponding to natural allelic variants and homologues of the VEGF-G cDNAs of the invention can be isolated based on their homology to the VEGF-G nucleic acids disclosed herein using the cDNAs disclosed herein, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. Nucleic acid molecules corresponding to natural allelic variants and homologues of the VEGF-G cDNAs of the invention can further be isolated by mapping to the same chromosome or locus as the VEGF-G gene.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 15, 20, 25, 30 or more nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, 3, 10 or 12. In other embodiment, the nucleic acid is at least 30, 50, 100, 150, 200, 250, 253, 300, 350, 400, 450, 500, 549, 550, 600, 650, 700, 750, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200 or 2273 nucleotides in length. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other. Preferably, the conditions are such that sequences at least about 70%, more preferably at least about 80%, even more preferably at least about 85% or 90% homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 50°C. Another example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 55°C. A further example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 60°C. Preferably, stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X

SSC, 0.1% SDS at 65°C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO:1, 3, 10 or 12 corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

In addition to naturally-occurring allelic variants of the VEGF-G sequences that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequences of SEQ ID NO:1, 3, 10 or 12, thereby leading to changes in the amino acid sequence of the encoded VEGF-G proteins, without altering the functional ability of the VEGF-G proteins. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of SEQ ID NO:1, 3, 10 or 12. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of VEGF-G (*e.g.*, the sequence of SEQ ID NO:2 or 11) without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For example, amino acid residues that are conserved among the VEGF-G proteins of the present invention, *e.g.*, those present in the VEGF/PDGF superfamily variant motif, the VEGF-G disulfide knot-like domain, or the CUB domain, are predicted to be particularly unamenable to alteration. Furthermore, additional amino acid residues that are conserved between the VEGF-G proteins of the present invention and other members of the VEGF family are not likely to be amenable to alteration.

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding VEGF-G proteins that contain changes in amino acid residues that are not essential for activity. Such VEGF-G proteins differ in amino acid sequence from SEQ ID NO:2 or 11, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or more homologous to SEQ ID NO:2 or 11.

An isolated nucleic acid molecule encoding a VEGF-G protein homologous to the protein of SEQ ID NO:2 or 11 can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO:1, 3, 10

or 12 such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced into SEQ ID NO:1, 3, 10 or 12 by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in a VEGF-G protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a VEGF-G coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for VEGF-G biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO:1, 3, 10 or 12, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

In a preferred embodiment, a mutant VEGF-G protein can be assayed for the ability to (1) interact with a non-VEGF-G protein molecule, *e.g.*, a VEGF-G substrate or receptor; (2) activate a VEGF-G-dependent signal transduction pathway; (3) modulate cell proliferation, differentiation, migration and/or apoptosis mechanisms; or (4) modulate angiogenic processes.

In addition to the nucleic acid molecules encoding VEGF-G proteins described above, another aspect of the invention pertains to isolated nucleic acid molecules which are antisense thereto. An "antisense" nucleic acid comprises a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or

complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire VEGF-G coding strand, or to only a portion thereof. In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding VEGF-G. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (*e.g.*, the coding region of human VEGF-G corresponds to SEQ ID NO:3 and the coding region of mouse VEGF-G corresponds to SEQ ID NO:12). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding VEGF-G. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding VEGF-G disclosed herein (*e.g.*, SEQ ID NO:3 or 12), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of VEGF-G mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of VEGF-G mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of VEGF-G mRNA. An antisense oligonucleotide can be, for example, about 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, or more nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-

carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5- oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a VEGF-G protein to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention include direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense

nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids. Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue *et al.* (1987) *FEBS Lett.* 215:327-330).

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave VEGF-G mRNA transcripts to thereby inhibit translation of VEGF-G mRNA. A ribozyme having specificity for a VEGF-G-encoding nucleic acid can be designed based upon the nucleotide sequence of a VEGF-G cDNA disclosed herein (*i.e.*, SEQ ID NO:1, 3, 10 or 12). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a VEGF-G-encoding mRNA. See, *e.g.*, Cech *et al.* U.S. Patent No. 4,987,071; and Cech *et al.* U.S. Patent No. 5,116,742. Alternatively, VEGF-G mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel, D. and Szostak, J.W. (1993) *Science* 261:1411-1418.

Alternatively, VEGF-G gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the VEGF-G (*e.g.*, the VEGF-G promoter and/or enhancers) to form triple helical structures that prevent transcription of the VEGF-G gene in target cells. See generally, Helene, C. (1991) *Anticancer Drug Des.* 6(6):569-84; Helene, C. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher, L.J. (1992) *Bioassays* 14(12):807-15.

In yet another embodiment, the VEGF-G nucleic acid molecules of the present invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acid molecules can be modified to generate peptide nucleic acids (see Hyrup B. *et al.* (1996) *Bioorganic & Medicinal Chemistry* 4 (1): 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup B. *et al.* (1996) *supra*; Perry-O'Keefe *et al.* Proc. Natl. Acad. Sci. 93: 14670-675.

PNAs of VEGF-G nucleic acid molecules can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, for example, inducing transcription or translation arrest or inhibiting replication. PNAs of VEGF-G nucleic acid molecules can also be used in the analysis of single base pair mutations in a gene, (*e.g.*, by PNA-directed PCR clamping); as 'artificial restriction enzymes' when used in combination with other enzymes, (*e.g.*, S1 nucleases (Hyrup B. (1996) *supra*)); or as probes or primers for DNA sequencing or hybridization (Hyrup B. *et al.* (1996) *supra*; Perry-O'Keefe *supra*).

In another embodiment, PNAs of VEGF-G can be modified, (*e.g.*, to enhance their stability or cellular uptake), by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of VEGF-G nucleic acid molecules can be generated which may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, (*e.g.*, RNAse H and DNA polymerases), to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of

bonds between the nucleobases, and orientation (Hyrup B. (1996) *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup B. (1996) *supra* and Finn P.J. *et al.* (1996) *Nucleic Acids Res.* 24 (17): 3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used as a between the PNA and the 5' end of DNA (Mag, M. *et al.* (1989) *Nucleic Acid Res.* 17: 5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn P.J. *et al.* (1996) *supra*). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment (Peterser, K.H. *et al.* (1975) *Bioorganic Med. Chem. Lett.* 5: 1119-11124).

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:6553-6556; Lemaitre *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, *e.g.*, PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (See, *e.g.*, Krol *et al.* (1988) *Bio-Techniques* 6:958-976) or intercalating agents. (See, *e.g.*, Zon (1988) *Pharm. Res.* 5:539-549). To this end, the oligonucleotide may be conjugated to another molecule, (*e.g.*, a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent).

II. Isolated VEGF-G Proteins and Anti-VEGF-G Antibodies

One aspect of the invention pertains to isolated VEGF-G proteins, and biologically active portions thereof, as well as polypeptide fragments suitable for use as immunogens to raise anti-VEGF-G antibodies. In one embodiment, native VEGF-G proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, VEGF-G proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a VEGF-G protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the VEGF-G protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of VEGF-G protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of VEGF-G protein having less than about 30% (by dry weight) of non-VEGF-G protein (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-VEGF-G protein, still more preferably less than about 10% of non-VEGF-G protein, and most preferably less than about 5% non-VEGF-G protein. When the VEGF-G protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of VEGF-G protein in which the protein is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of VEGF-G protein having less than about 30% (by dry weight) of chemical precursors or non-VEGF-G chemicals, more preferably less than about 20% chemical precursors or non-VEGF-G chemicals, still more preferably less than about 10% chemical precursors or non-VEGF-G chemicals, and most preferably less than about 5% chemical precursors or non-VEGF-G chemicals.

As used herein, a "biologically active portion" of a VEGF-G protein includes a fragment of a VEGF-G protein which participates in an interaction between a VEGF-G molecule and a non-VEGF-G molecule. Biologically active portions of a VEGF-G protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of the VEGF-G protein, *e.g.*, the amino acid sequence shown in SEQ ID NO:2 or 11, which include less amino acids than the full

length VEGF-G proteins, and exhibit at least one activity of a VEGF-G protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the VEGF-G protein, *e.g.*, modulating cell growth and/or migration mechanisms. A biologically active portion of a VEGF-G protein can be a polypeptide which is, for example, 10, 25, 50, 100, 200 or more amino acids in length. Biologically active portions of a VEGF-G protein can be used as targets for developing agents which modulate a VEGF-G mediated activity, *e.g.*, a cell proliferation, differentiation, migration, apoptosis, or angiogenic signalling mechanism.

In one embodiment, a biologically active portion of a VEGF-G protein comprises at least one VEGF/PDGF superfamily variant motif, and/or at least one CUB domain. It is to be understood that a preferred biologically active portion of a VEGF-G protein of the present invention may contain at least one VEGF/PDGF superfamily variant motif. In another embodiment, a biologically active portion of a VEGF-G protein comprises at least one VEGF-G disulfide knot-like domain, and/or at least one CUB domain. It is to be understood that a preferred biologically active portion of a VEGF-G protein of the present invention may contain at least one VEGF-G disulfide knot-like domain. Another preferred biologically active portion of a VEGF-G protein may contain at least one CUB domain. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native VEGF-G protein.

In a preferred embodiment, the VEGF-G protein has an amino acid sequence shown in SEQ ID NO:2 or 11. In other embodiments, the VEGF-G protein is substantially homologous to SEQ ID NO:2 or 11, and retains the functional activity of the protein of SEQ ID NO:2 or 11, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail in subsection I above. Accordingly, in another embodiment, the VEGF-G protein is a protein which comprises an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or more homologous to SEQ ID NO:2 or 11.

To determine the percent identity of two amino acid sequences or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 80%, or 90% of the length of the reference sequence (*e.g.*, when aligning a second sequence to the VEGF-G amino acid sequence of SEQ ID NO:2 or 11 having 370 amino acid residues, at least 111, preferably at least 148, more preferably at least 185, even more preferably at least 222, and even more preferably at least 259, 296 or 333 amino acid residues are aligned). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package (available at http://www.gcg.com), using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (available at http://www.gcg.com), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent

identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Meyers and W. Miller (*Comput. Appl. Biosci.*, 4:11-17 (1988)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences of the present invention can further be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, *et al.* (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to VEGF-G nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to VEGF-G protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.*, (1997) *Nucleic Acids Res.* 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used. See http://www.ncbi.nlm.nih.gov.

The invention also provides VEGF-G chimeric or fusion proteins. As used herein, a VEGF-G "chimeric protein" or "fusion protein" comprises a VEGF-G polypeptide operatively linked to a non-VEGF-G polypeptide. An "VEGF-G polypeptide" refers to a polypeptide having an amino acid sequence corresponding to VEGF-G, whereas a "non-VEGF-G polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the VEGF-G protein, *e.g.*, a protein which is different from the VEGF-G protein and which is derived from the same or a different organism. Within a VEGF-G fusion protein the VEGF-G polypeptide can correspond to all or a portion of a VEGF-G protein. In a preferred embodiment, a VEGF-G fusion protein comprises at least one biologically active portion of a VEGF-G protein. In another preferred embodiment, a VEGF-G fusion protein comprises at least two biologically active portions of a VEGF-G protein. Within the fusion protein, the term "operatively linked" is intended to indicate

that the VEGF-G polypeptide and the non-VEGF-G polypeptide are fused in-frame to each other. The non-VEGF-G polypeptide can be fused to the N-terminus or C-terminus of the VEGF-G polypeptide.

For example, in one embodiment, the fusion protein is a GST-VEGF-G fusion protein in which the VEGF-G sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant VEGF-G.

In another embodiment, the fusion protein is an alkaline phosphatase-VEGF-G fusion protein in which the VEGF-G sequences are fused to alkaline phosphatase sequences. Such fusion proteins can be used to assess VEGF-G binding to cells and/or tissues.

In another embodiment, the fusion protein is a VEGF-G protein containing a heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of VEGF-G can be increased through use of a heterologous signal sequence.

The VEGF-G fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject *in vivo*. The VEGF-G fusion proteins can be used to affect the bioavailability of a VEGF-G substrate. Use of VEGF-G fusion proteins may be useful therapeutically for the treatment of disorders caused by, for example, (i) aberrant modification or mutation of a gene encoding a VEGF-G protein; (ii) mis-regulation of the VEGF-G gene; and (iii) aberrant post-translational modification of a VEGF-G protein.

Moreover, the VEGF-G-fusion proteins of the invention can be used as immunogens to produce anti-VEGF-G antibodies in a subject, to purify VEGF-G ligands and in screening assays to identify molecules which inhibit the interaction of VEGF-G with a VEGF-G substrate.

Preferably, a VEGF-G chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid

undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel *et al.* John Wiley & Sons: 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A VEGF-G-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the VEGF-G protein.

The present invention also pertains to variants of the VEGF-G proteins which function as either VEGF-G agonists (mimetics) or as VEGF-G antagonists. Variants of the VEGF-G proteins can be generated by mutagenesis, *e.g.*, discrete point mutation or truncation of a VEGF-G protein. An agonist of the VEGF-G proteins can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of a VEGF-G protein. An antagonist of a VEGF-G protein can inhibit one or more of the activities of the naturally occurring form of the VEGF-G protein by, for example, competitively modulating a VEGF-G-mediated activity of a VEGF-G protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the VEGF-G protein.

In one embodiment, variants of a VEGF-G protein which function as either VEGF-G agonists (mimetics) or as VEGF-G antagonists can be identified by screening combinatorial libraries of mutants, *e.g.*, truncation mutants, of a VEGF-G protein for VEGF-G protein agonist or antagonist activity. In one embodiment, a variegated library of VEGF-G variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of VEGF-G variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential VEGF-G

sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of VEGF-G sequences therein. There are a variety of methods which can be used to produce libraries of potential VEGF-G variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential VEGF-G sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, *e.g.*, Narang, S.A. (1983) *Tetrahedron* 39:3; Itakura *et al.* (1984) *Annu. Rev. Biochem.* 53:323; Itakura *et al.* (1984) *Science* 198:1056; Ike *et al.* (1983) *Nucleic Acid Res.* 11:477.

In addition, libraries of fragments of a VEGF-G protein coding sequence can be used to generate a variegated population of VEGF-G fragments for screening and subsequent selection of variants of a VEGF-G protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a VEGF-G coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal, C-terminal and internal fragments of various sizes of the VEGF-G protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of VEGF-G proteins. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in

which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recrusive ensemble mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify VEGF-G variants (Arkin and Yourvan (1992) *Proc. Natl. Acad. Sci. USA* 89:7811-7815; Delgrave *et al.* (1993) *Protein Engineering* 6(3):327-331).

In one embodiment, cell based assays can be exploited to analyze a variegated VEGF-G library. For example, a library of expression vectors can be transfected into a cell line, *e.g.*, an endothelial cell line, which ordinarily responds to VEGF-G in a particular VEGF-G substrate-dependent manner. The transfected cells are then contacted with VEGF-G and the effect of the expression of the mutant on signalling by the VEGF-G substrate can be detected, *e.g.*, by measuring intracellular calcium and inositol 1,4,5-trisphosphate (IP3) levels, cell growth, and cell migration. Plasmid DNA can then be recovered from the cells which score for inhibition, or alternatively, potentiation of signalling by the VEGF-G substrate, and the individual clones further characterized.

An isolated VEGF-G protein, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind VEGF-G using standard techniques for polyclonal and monoclonal antibody preparation. A full-length VEGF-G protein can be used or, alternatively, the invention provides antigenic peptide fragments of VEGF-G for use as immunogens. The antigenic peptide of VEGF-G comprises at least 8 amino acid residues of the amino acid sequence shown in SEQ ID NO:2 and encompasses an epitope of VEGF-G such that an antibody raised against the peptide forms a specific immune complex with VEGF-G. Preferably, the antigenic peptide comprises at least 10 amino acid residues, more preferably at least 15 amino acid residues, even more preferably at least 20 amino acid residues, and most preferably at least 30 amino acid residues.

Preferred epitopes encompassed by the antigenic peptide are regions of VEGF-G that are located on the surface of the protein, *e.g.*, hydrophilic regions, as well as regions with high antigenicity.

A VEGF-G immunogen typically is used to prepare antibodies by immunizing a suitable subject, (*e.g.*, rabbit, goat, mouse or other mammal) with the immunogen. An appropriate immunogenic preparation can contain, for example, recombinantly expressed VEGF-G protein or a chemically synthesized VEGF-G polypeptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or similar immunostimulatory agent. Immunization of a suitable subject with an immunogenic VEGF-G preparation induces a polyclonal anti-VEGF-G antibody response.

Accordingly, another aspect of the invention pertains to anti-VEGF-G antibodies. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, *i.e.*, molecules that contain an antigen binding site which specifically binds (immunoreacts with) an antigen, such as VEGF-G. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and $F(ab')_2$ fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies that bind VEGF-G. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of VEGF-G. A monoclonal antibody composition thus typically displays a single binding affinity for a particular VEGF-G protein with which it immunoreacts.

Polyclonal anti-VEGF-G antibodies can be prepared as described above by immunizing a suitable subject with a VEGF-G immunogen. The anti-VEGF-G antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized VEGF-G. If desired, the antibody molecules directed against VEGF-G can be isolated from the mammal (*e.g.*, from the blood) and further purified by well known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after immunization, *e.g.*, when the anti-VEGF-G antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described

by Kohler and Milstein (1975) *Nature* 256:495-497) (see also, Brown *et al.* (1981) *J. Immunol.* 127:539-46; Brown *et al.* (1980) *J. Biol. Chem* .255:4980-83; Yeh *et al.* (1976) *Proc. Natl. Acad. Sci. USA* 76:2927-31; and Yeh *et al.* (1982) *Int. J. Cancer* 29:269-75), the more recent human B cell hybridoma technique (Kozbor *et al.* (1983) *Immunol Today* 4:72), the EBV-hybridoma technique (Cole *et al.* (1985), *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for producing monoclonal antibody hybridomas is well known (see generally R. H. Kenneth, in *Monoclonal Antibodies: A New Dimension In Biological Analyses*, Plenum Publishing Corp., New York, New York (1980); E. A. Lerner (1981) *Yale J. Biol. Med.*, 54:387-402; M. L. Gefter *et al.* (1977) *Somatic Cell Genet.* 3:231-36). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized with a VEGF-G immunogen as described above, and the culture supernatants of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds VEGF-G.

Any of the many well known protocols used for fusing lymphocytes and immortalized cell lines can be applied for the purpose of generating an anti-VEGF-G monoclonal antibody (see, *e.g.*, G. Galfre *et al.* (1977) *Nature* 266:55052; Gefter *et al.* *Somatic Cell Genet.*, cited *supra*; Lerner, *Yale J. Biol. Med.*, cited *supra*; Kenneth, *Monoclonal Antibodies*, cited *supra*). Moreover, the ordinarily skilled worker will appreciate that there are many variations of such methods which also would be useful. Typically, the immortal cell line (*e.g.*, a myeloma cell line) is derived from the same mammalian species as the lymphocytes. For example, murine hybridomas can be made by fusing lymphocytes from a mouse immunized with an immunogenic preparation of the present invention with an immortalized mouse cell line. Preferred immortal cell lines are mouse myeloma cell lines that are sensitive to culture medium containing hypoxanthine, aminopterin and thymidine ("HAT medium"). Any of a number of myeloma cell lines can be used as a fusion partner according to standard techniques, *e.g.*, the P3-NS1/1-Ag4-1, P3-x63-Ag8.653 or Sp2/O-Ag14 myeloma lines. These myeloma lines are available from ATCC. Typically, HAT-sensitive mouse myeloma cells are fused to mouse splenocytes using polyethylene glycol ("PEG"). Hybridoma cells resulting from the fusion are then selected using HAT medium, which kills unfused

and unproductively fused myeloma cells (unfused splenocytes die after several days because they are not transformed). Hybridoma cells producing a monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind VEGF-G, *e.g.*, using a standard ELISA assay.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal anti-VEGF-G antibody can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (*e.g.*, an antibody phage display library) with VEGF-G to thereby isolate immunoglobulin library members that bind VEGF-G. Kits for generating and screening phage display libraries are commercially available (*e.g.*, the Pharmacia *Recombinant Phage Antibody System*, Catalog No. 27-9400-01; and the Stratagene *SurfZAP™ Phage Display Kit*, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, Ladner *et al.* U.S. Patent No. 5,223,409; Kang *et al.* PCT International Publication No. WO 92/18619; Dower *et al.* PCT International Publication No. WO 91/17271; Winter *et al.* PCT International Publication WO 92/20791; Markland *et al.* PCT International Publication No. WO 92/15679; Breitling *et al.* PCT International Publication WO 93/01288; McCafferty *et al.* PCT International Publication No. WO 92/01047; Garrard *et al.* PCT International Publication No. WO 92/09690; Ladner *et al.* PCT International Publication No. WO 90/02809; Fuchs *et al.* (1991) *Bio/Technology* 9:1370-1372; Hay *et al.* (1992) *Hum. Antibod. Hybridomas* 3:81-85; Huse *et al.* (1989) *Science* 246:1275-1281; Griffiths *et al.* (1993) *EMBO J* 12:725-734; Hawkins *et al.* (1992) *J. Mol. Biol.* 226:889-896; Clarkson *et al.* (1991) *Nature* 352:624-628; Gram *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:3576-3580; Garrad *et al.* (1991) *Bio/Technology* 9:1373-1377; Hoogenboom *et al.* (1991) *Nuc. Acid Res.* 19:4133-4137; Barbas *et al.* (1991) *Proc. Natl. Acad. Sci. USA* 88:7978-7982; and McCafferty *et al. Nature* (1990) 348:552-554.

Additionally, recombinant anti-VEGF-G antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in

Robinson *et al.* International Application No. PCT/US86/02269; Akira, *et al.* European Patent Application 184,187; Taniguchi, M., European Patent Application 171,496; Morrison *et al.* European Patent Application 173,494; Neuberger *et al.* PCT International Publication No. WO 86/01533; Cabilly *et al.* U.S. Patent No. 4,816,567; Cabilly *et al.* European Patent Application 125,023; Better *et al.* (1988) *Science* 240:1041-1043; Liu *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:3439-3443; Liu *et al.* (1987) *J. Immunol.* 139:3521-3526; Sun *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:214-218; Nishimura *et al.* (1987) *Canc. Res.* 47:999-1005; Wood *et al.* (1985) *Nature* 314:446-449; and Shaw *et al.* (1988) *J. Natl. Cancer Inst.* 80:1553-1559); Morrison, S. L. (1985) *Science* 229:1202-1207; Oi *et al.* (1986) *BioTechniques* 4:214; Winter U.S. Patent 5,225,539; Jones *et al.* (1986) *Nature* 321:552-525; Verhoeyan *et al.* (1988) *Science* 239:1534; and Beidler *et al.* (1988) *J. Immunol.* 141:4053-4060.

An anti-VEGF-G antibody (*e.g.*, monoclonal antibody) can be used to isolate VEGF-G by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-VEGF-G antibody can facilitate the purification of natural VEGF-G from cells and of recombinantly produced VEGF-G expressed in host cells. Moreover, an anti-VEGF-G antibody can be used to detect VEGF-G protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the VEGF-G protein. Anti-VEGF-G antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent

materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

III. Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a VEGF-G protein (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a

host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cells and those which direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, and the like. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, VEGF-G proteins, mutant forms of VEGF-G proteins, fusion proteins, and the like).

The recombinant expression vectors of the invention can be designed for expression of VEGF-G proteins in prokaryotic or eukaryotic cells. For example, VEGF-G proteins can be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein

from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D.B. and Johnson, K.S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Purified fusion proteins can be utilized in VEGF-G activity assays, (*e.g.*, direct assays or competitive assays described in detail below), or to generate antibodies specific for VEGF-G proteins, for example. In a preferred embodiment, a VEGF-G fusion protein expressed in a retroviral expression vector of the present invention can be utilized to infect bone marrow cells which are subsequently transplanted into irradiated recipients. The pathology of the subject recipient is then examined after sufficient time has passed (*e.g.*, six (6) weeks).

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada *et al.*, (1992) *Nucleic Acids Res.* 20:2111-2118).

Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the VEGF-G expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerivisae* include pYepSec1 (Baldari, *et al.*, (1987) *Embo J.* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz *et al.*, (1987) *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, CA), and picZ (InVitrogen Corp, San Diego, CA).

Alternatively, VEGF-G proteins can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, Sf 9 cells) include the pAc series (Smith *et al.* (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, B. (1987) *Nature* 329:840) and pMT2PC (Kaufman *et al.* (1987) *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert *et al.* (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji *et al.* (1983) *Cell*

33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle (1989) *Proc. Natl. Acad. Sci. USA* 86:5473-5477), endothelial cell-specific promoters (*e.g.*, KDR/flk promoter; U.S. Patent No. 5,888,765), pancreas-specific promoters (Edlund *et al.* (1985) *Science* 230:912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α-fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to VEGF-G mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H. *et al.*, Antisense RNA as a molecular tool for genetic analysis, *Reviews - Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a VEGF-G nucleic acid molecule of the invention is introduced, *e.g.*, a VEGF-G nucleic acid molecule within a recombinant expression vector or a VEGF-G nucleic acid molecule containing sequences which allow it to homologously recombine into a specific site of the host cell's genome. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular

subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, a VEGF-G protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as human umbilical vein endothelial cells (HUVEC), human microvascular endothelial cells (HMVEC), Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding a VEGF-G protein or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) a VEGF-G protein. Accordingly, the invention further provides methods for producing a VEGF-G protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of the invention (into which a recombinant expression vector encoding a VEGF-G protein has been introduced) in a suitable medium such that a VEGF-G protein is produced. In another embodiment, the method further comprises isolating a VEGF-G protein from the medium or the host cell.

The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which VEGF-G-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous VEGF-G sequences have been introduced into their genome or homologous recombinant animals in which endogenous VEGF-G sequences have been altered. Such animals are useful for studying the function and/or activity of a VEGF-G protein and for identifying and/or evaluating modulators of VEGF-G activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, and the like. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous VEGF-G gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing a VEGF-G-encoding nucleic acid into the male pronuclei of a fertilized oocyte, *e.g.*, by microinjection, retroviral infection, and allowing the oocyte to develop in a

pseudopregnant female foster animal. The VEGF-G cDNA sequence of SEQ ID NO:1 or 10 can be introduced as a transgene into the genome of a non-human animal. Alternatively, a non-human or non-mouse homologue of a human or mouse VEGF-G gene, such as a rat or VEGF-G gene, can be used as a transgene. Alternatively, a VEGF-G gene homologue, such as another VEGF-G family member, can be isolated based on hybridization to the VEGF-G cDNA sequences of SEQ ID NO:1, 3, 10 or 12 (described further in subsection I above) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to a VEGF-G transgene to direct expression of a VEGF-G protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of a VEGF-G transgene in its genome and/or expression of VEGF-G mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding a VEGF-G protein can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of a VEGF-G gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the VEGF-G gene. The VEGF-G gene can be a human gene (*e.g.*, the cDNA of SEQ ID NO:3), but more preferably, is a non-human homolog of a human VEGF-G gene (*e.g.*, a cDNA isolated by stringent hybridization with the nucleotide sequence of SEQ ID NO:1 or 10), For example, a mouse VEGF-G gene (*e.g.*, the cDNA of SEQ ID NO:12) can be used to construct a homologous recombination nucleic acid molecule, *e.g.*, a vector, suitable for altering an endogenous VEGF-G gene in the mouse genome. In a preferred

embodiment, the homologous recombination nucleic acid molecule is designed such that, upon homologous recombination, the endogenous VEGF-G gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the homologous recombination nucleic acid molecule can be designed such that, upon homologous recombination, the endogenous VEGF-G gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous VEGF-G protein). In the homologous recombination nucleic acid molecule, the altered portion of the VEGF-G gene is flanked at its 5' and 3' ends by additional nucleic acid sequence of the VEGF-G gene to allow for homologous recombination to occur between the exogenous VEGF-G gene carried by the homologous recombination nucleic acid molecule and an endogenous VEGF-G gene in a cell, *e.g.*, an embryonic stem cell. The additional flanking VEGF-G nucleic acid sequence is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the homologous recombination nucleic acid molecule (see, *e.g.*, Thomas, K.R. and Capecchi, M. R. (1987) *Cell* 51:503 for a description of homologous recombination vectors). The homologous recombination nucleic acid molecule is introduced into a cell, *e.g.*, an embryonic stem cell line (*e.g.*, by electroporation) and cells in which the introduced VEGF-G gene has homologously recombined with the endogenous VEGF-G gene are selected (see *e.g.*, Li, E. *et al.* (1992) *Cell* 69:915). The selected cells can then injected into a blastocyst of an animal (*e.g.*, a mouse) to form aggregation chimeras (see *e.g.*, Bradley, A. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E.J. Robertson, ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination nucleic acid molecules, *e.g.*, vectors, or homologous recombinant animals are described further in Bradley, A. (1991) *Current Opinion in Biotechnology* 2:823-829 and in PCT International Publication Nos.: WO 90/11354 by Le Mouellec *et*

al.; WO 91/01140 by Smithies *et al.*; WO 92/0968 by Zijlstra *et al.*; and WO 93/04169 by Berns *et al.*

In another embodiment, transgenic non-humans animals can be produced which contain selected systems which allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, *e.g.*, Lakso *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman *et al.* (1991) *Science* 251:1351-1355. If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, *e.g.*, by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, I. *et al.* (1997) *Nature* 385:810-813 and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, *e.g.*, a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G_0 phase. The quiescent cell can then be fused, *e.g.*, through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell, *e.g.*, the somatic cell, is isolated.

IV. Pharmaceutical Compositions

The VEGF-G nucleic acid molecules, fragments of VEGF-G proteins, and anti-VEGF-G antibodies (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions suitable for administration. Active compounds may include, but are not limited to, peptides, nucleic acids, antibodies, and small inorganic or inorganic compounds. Such compositions typically comprise the

nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as

bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyetheylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, a fragment of a VEGF-G protein or an anti-VEGF-G antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as

microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD50 (the dose lethal to 50% of the population) and the ED50 (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD50/ED50. Compounds which exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC50 (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

As defined herein, a therapeutically effective amount of protein or polypeptide (*i.e.*, an effective dosage) ranges from about 0.001 to 30 mg/kg body weight, preferably about 0.01 to 25 mg/kg body weight, more preferably about 0.1 to 20 mg/kg body weight, and even more preferably about 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7 mg/kg, or 5 to 6 mg/kg body weight. The skilled artisan will appreciate that certain factors may influence the dosage required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a protein, polypeptide, or antibody can include a single treatment or, preferably, can include a series of treatments.

In a preferred example, a subject is treated with antibody, protein, or polypeptide in the range of between about 0.1 to 20 mg/kg body weight, one time per week for between about 1 to 10 weeks, preferably between 2 to 8 weeks, more preferably between about 3 to 7 weeks, and even more preferably for about 4, 5, or 6 weeks. It will also be appreciated that the effective dosage of antibody, protein, or polypeptide used for treatment may increase or decrease over the course of a particular treatment. Changes in dosage may result and become apparent from the results of diagnostic assays as described herein.

The present invention encompasses agents which modulate expression or activity. An agent may, for example, be a small molecule. For example, such small molecules include, but are not limited to, peptides, peptidomimetics, amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (i.e., including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds.

It is understood that appropriate doses of small molecule agents depends upon a number of factors within the ken of the ordinarily skilled physician, veterinarian, or researcher. The dose(s) of the small molecule will vary, for example, depending upon the identity, size, and condition of the subject or sample being treated, further depending upon the route by which the composition is to be administered, if applicable, and the effect which the practitioner desires the small molecule to have upon the nucleic acid or polypeptide of the invention.

Exemplary doses include milligram or microgram amounts of the small molecule per kilogram of subject or sample weight (*e.g.*, about 1microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1microgram per kilogram to about 50 micrograms per kilogram. It is furthermore understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. Such appropriate doses may be determined using the assays described herein. When one or more of these small molecules is to be administered to an animal (*e.g.*, a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

Further, an antibody (or fragment thereof) may be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive metal ion. A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1-dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, and puromycin and analogs

or homologs thereof. Therapeutic agents include, but are not limited to, antimetabolites (*e.g.*, methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (*e.g.*, mechlorethamine, thioepa chlorambucil, melphalan, carmustine (BSNU) and lomustine (CCNU), cyclothosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (*e.g.*, daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (*e.g.*, dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (*e.g.*, vincristine and vinblastine).

The conjugates of the invention can be used for modifying a given biological response, the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a protein such as tumor necrosis factor, .alpha.-interferon, .beta.-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophage colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors.

Techniques for conjugating such therapeutic moiety to antibodies are well known, see, *e.g.*, Arnon *et al.*, "Monoclonal Antibodies For Immunotargeting Of Drugs In Cancer Therapy", in Monoclonal Antibodies And Cancer Therapy, Reisfeld *et al.* (eds.), pp. 243-56 (Alan R. Liss, Inc. 1985); Hellstrom *et al.*, "Antibodies For Drug Delivery", in Controlled Drug Delivery (2nd Ed.), Robinson *et al.* (eds.), pp. 623-53 (Marcel Dekker, Inc. 1987); Thorpe, "Antibody Carriers Of Cytotoxic Agents In Cancer Therapy: A Review", in Monoclonal Antibodies '84: Biological And Clinical Applications, Pinchera *et al.* (eds.), pp. 475-506 (1985); "Analysis, Results, And Future Prospective Of The Therapeutic Use Of Radiolabeled Antibody In Cancer Therapy", in Monoclonal Antibodies For Cancer Detection And Therapy, Baldwin *et al.* (eds.), pp. 303-16 (Academic Press 1985), and Thorpe *et al.*, "The Preparation And Cytotoxic Properties Of Antibody-Toxin Conjugates".

Immunol. Rev., 62:119-58 (1982). Alternatively, an antibody can be conjugated to a second antibody to Form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (see U.S. Patent 5,328,470) or by stereotactic injection (see *e.g.*, Chen *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

V. Uses and Methods of the Invention

The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: a) screening assays; b) predictive medicine (*e.g.*, diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenetics); and c) methods of treatment (*e.g.*, therapeutic and prophylactic). As described herein, a VEGF-G protein of the invention has one or more of the following activities: (1) it interacts with a non-VEGF-G protein molecule, *e.g.*, a VEGF-G substrate, such as a VEGF receptor; (2) it activates a VEGF-G-dependent signal transduction pathway; (3) it modulates cell proliferation, differentiation, and/or migration mechanisms; (4) it modulates angiogenesis, and, thus, can be used to, for example, (1) modulate the interaction with a non-VEGF-G protein molecule; (2) to activate a VEGF-G-dependent signal transduction pathway; (3) to modulate cell proliferation, differentiation, and/or migration mechanisms; (4) to modulate angiogenesis.

The isolated nucleic acid molecules of the invention can be used, for example, to express VEGF-G protein (*e.g.*, via a recombinant expression vector in a host cell in gene therapy applications), to detect VEGF-G mRNA (*e.g.*, in a biological sample) or a genetic alteration in a VEGF-G gene, and to modulate VEGF-G activity, as described further below. The VEGF-G proteins can be used to treat disorders characterized by insufficient or excessive production of a VEGF-G substrate or production of VEGF-G inhibitors. In addition, the VEGF-G proteins can be used to screen for naturally occurring VEGF-G substrates, to screen for drugs or compounds which modulate VEGF-G activity, as well as to treat disorders characterized by insufficient or excessive production of VEGF-G protein or production of VEGF-G protein forms which have decreased, aberrant or unwanted activity compared to VEGF-G wild type protein (*e.g.*, cell proliferation and/or differentiation disorders, such as disorders characterized by aberrant angiogenesis). Moreover, the anti-VEGF-G antibodies of the invention can be used to detect and isolate VEGF-G proteins, regulate the bioavailability of VEGF-G proteins, and modulate VEGF-G activity.

A. Screening Assays:

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) which bind to VEGF-G proteins, have a stimulatory or inhibitory effect on, for example, VEGF-G expression or VEGF-G activity, or have a stimulatory or inhibitory effect on, for example, the expression or activity of a VEGF-G substrate.

In one embodiment, the invention provides assays for screening candidate or test compounds which are substrates of a VEGF-G protein or polypeptide or biologically active portion thereof. In another embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of a VEGF-G protein or polypeptide or biologically active portion thereof. The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic

library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K.S. (1997) *Anticancer Drug Des.* 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:6909; Erb *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann *et al.* (1994). *J. Med. Chem.* 37:2678; Cho *et al.* (1993) *Science* 261:1303; Carrell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and in Gallop *et al.* (1994) *J. Med. Chem.* 37:1233.

Libraries of compounds may be presented in solution (*e.g.*, Houghten (1992) *Biotechniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria (Ladner USP 5,223,409), spores (Ladner USP '409), plasmids (Cull *et al.* (1992) *Proc Natl Acad Sci USA* 89:1865-1869) or on phage (Scott and Smith (1990) *Science* 249:386-390); (Devlin (1990) *Science* 249:404-406); (Cwirla *et al.* (1990) *Proc. Natl. Acad. Sci.* 87:6378-6382); (Felici (1991) *J. Mol. Biol.* 222:301-310); (Ladner *supra.*).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a VEGF-G protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to modulate VEGF-G activity is determined. Determining the ability of the test compound to modulate VEGF-G activity can be accomplished by monitoring, for example, intracellular calcium and inositol 1,4,5-trisphosphate (IP3) levels, cell growth, and cell chemotaxis. The cell, for example, can be of mammalian origin, *e.g.*, an endothelial cell.

The ability of the test compound to modulate VEGF-G binding to a substrate or to bind to VEGF-G can also be determined. Determining the ability of the test compound to modulate VEGF-G binding to a substrate can be accomplished, for example, by coupling the VEGF-G substrate with a radioisotope or enzymatic label such that binding of the VEGF-G substrate to VEGF-G can be determined by detecting the labeled VEGF-G substrate in a complex. Alternatively, VEGF-G could be coupled with

a radioisotope or enzymatic label to monitor the ability of a test compound to modulate VEGF-G binding to a VEGF-G substrate in a complex. Determining the ability of the test compound to bind VEGF-G can be accomplished, for example, by coupling the compound with a radioisotope or enzymatic label such that binding of the compound to VEGF-G can be determined by detecting the labeled VEGF-G compound in a complex. For example, compounds (*e.g.*, VEGF-G substrates) can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product.

It is also within the scope of this invention to determine the ability of a compound (*e.g.*, a VEGF-G substrate) to interact with VEGF-G without the labeling of any of the interactants. For example, a microphysiometer can be used to detect the interaction of a compound with VEGF-G without the labeling of either the compound or the VEGF-G. McConnell, H. M. *et al.* (1992) *Science* 257:1906-1912. As used herein, a "microphysiometer" (*e.g.*, Cytosensor) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between a compound and VEGF-G.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a VEGF-G target molecule (*e.g.*, a VEGF-G substrate) with a test compound and determining the ability of the test compound to modulate (*e.g.* stimulate or inhibit) the activity of the VEGF-G target molecule. Determining the ability of the test compound to modulate the activity of a VEGF-G target molecule can be accomplished, for example, by determining the ability of the VEGF-G protein to bind to or interact with the VEGF-G target molecule.

Determining the ability of the VEGF-G protein or a biologically active fragment thereof, to bind to or interact with a VEGF-G target molecule can be accomplished by one of the methods described above for determining direct binding. In a preferred embodiment, determining the ability of the VEGF-G protein to bind to or interact with a

VEGF-G target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (*i.e.*, intracellular calcium or IP3), detecting catalytic/enzymatic activity of the target molecule upon an appropriate substrate, detecting the induction of a reporter gene (comprising a target-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, *e.g.*, luciferase), or detecting a target-regulated cellular response (*i.e.*, cell growth or migration).

In yet another embodiment, an assay of the present invention is a cell-free assay in which a VEGF-G protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to bind to the VEGF-G protein or biologically active portion thereof is determined. Preferred biologically active portions of the VEGF-G proteins to be used in assays of the present invention include fragments which participate in interactions with non-VEGF-G molecules, *e.g.*, fragments with high surface probability scores. Binding of the test compound to the VEGF-G protein can be determined either directly or indirectly as described above. In a preferred embodiment, the assay includes contacting the VEGF-G protein or biologically active portion thereof with a known compound which binds VEGF-G to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a VEGF-G protein, wherein determining the ability of the test compound to interact with a VEGF-G protein comprises determining the ability of the test compound to preferentially bind to VEGF-G or biologically active portion thereof as compared to the known compound.

In another embodiment, the assay is a cell-free assay in which a VEGF-G protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the VEGF-G protein or biologically active portion thereof is determined. Determining the ability of the test compound to modulate the activity of a VEGF-G protein can be accomplished, for example, by determining the ability of the VEGF-G protein to bind to a VEGF-G target molecule by one of the methods described above for determining direct binding. Determining the ability of the VEGF-G protein to bind to a VEGF-G target molecule

can also be accomplished using a technology such as real-time Biomolecular Interaction Analysis (BIA). Sjolander, S. and Urbaniczky, C. (1991) *Anal. Chem.* 63:2338-2345 and Szabo *et al.* (1995) *Curr. Opin. Struct. Biol.* 5:699-705. As used herein, "BIA" is a technology for studying biospecific interactions in real time, without labeling any of the interactants (*e.g.*, BIAcore). Changes in the optical phenomenon of surface plasmon resonance (SPR) can be used as an indication of real-time reactions between biological molecules.

In an alternative embodiment, determining the ability of the test compound to modulate the activity of a VEGF-G protein can be accomplished by determining the ability of the VEGF-G protein to further modulate the activity of a downstream effector of a VEGF-G target molecule. For example, the activity of the effector molecule on an appropriate target can be determined or the binding of the effector to an appropriate target can be determined as previously described.

In yet another embodiment, the cell-free assay involves contacting a VEGF-G protein or biologically active portion thereof with a known compound which binds the VEGF-G protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the VEGF-G protein, wherein determining the ability of the test compound to interact with the VEGF-G protein comprises determining the ability of the VEGF-G protein to preferentially bind to or modulate the activity of a VEGF-G target molecule.

The cell-free assays of the present invention are amenable to use of both soluble and/or membrane-bound forms of isolated proteins (*e.g.*, VEGF-G proteins or biologically active portions thereof). In the case of cell-free assays in which a membrane-bound form of an isolated protein is used it may be desirable to utilize a solubilizing agent such that the membrane-bound form of the isolated protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)$_n$, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-

cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl=N,N-dimethyl-3-ammonio-1-propane sulfonate.

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either VEGF-G or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to a VEGF-G protein, or interaction of a VEGF-G protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtitre plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase/VEGF-G fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or VEGF-G protein, and the mixture incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtitre plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of VEGF-G binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either a VEGF-G protein or a VEGF-G target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated VEGF-G protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques known in the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with VEGF-G protein or target molecules but which do not interfere with binding of the VEGF-G protein to its target molecule can be derivatized to the wells of the plate, and unbound target or

VEGF-G protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the VEGF-G protein or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the VEGF-G protein or target molecule.

In another embodiment, modulators of VEGF-G expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of VEGF-G mRNA or protein in the cell is determined. The level of expression of VEGF-G mRNA or protein in the presence of the candidate compound is compared to the level of expression of VEGF-G mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of VEGF-G expression based on this comparison. For example, when expression of VEGF-G mRNA or protein is greater (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of VEGF-G mRNA or protein expression. Alternatively, when expression of VEGF-G mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of VEGF-G mRNA or protein expression. The level of VEGF-G mRNA or protein expression in the cells can be determined by methods described herein for detecting VEGF-G mRNA or protein.

In yet another aspect of the invention, the VEGF-G proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, *e.g.*, U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell* 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.* 268:12046-12054; Bartel *et al.* (1993) *Biotechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with VEGF-G ("VEGF-G-binding proteins" or "VEGF-G-bp") and are involved in VEGF-G activity. Such VEGF-G-binding proteins are also likely to be involved in the propagation of signals by the VEGF-G proteins or VEGF-G targets as, for example, downstream elements of a VEGF-G-mediated signalling pathway. Alternatively, such VEGF-G-binding proteins are likely to be VEGF-G inhibitors.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a VEGF-G protein is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a VEGF-G-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the VEGF-G protein.

In another aspect, the invention pertains to a combination of two or more of the assays described herein. For example, a modulating agent can be identified using a cell-based or a cell free assay, and the ability of the agent to modulate the activity of a VEGF-G protein can be confirmed *in vivo*, *e.g.*, in an animal such as an animal model for angiogenesis, or for cellular transformation and/or tumorigenesis.

Cellular models for the study of angiogenesis include models of endothelial cell differentiation on Matrigel (Baatout, S. *et al.* (1996) *Rom. J. Intern. Med.* 34:263-269; Benelli, R *et al.* (1999) *Int. J. Biol. Markers* 14:243-246), embryonic stem cell models of vascular morphogenesis (Doetschman, T. *et al.* (1993) *Hypertension* 22:618-629), the culture of microvessel fragments in physiological gels (Hoying, JB *et al.* (1996) *In Vitro Cell Dev. Biol. Anim.* 32: 409-419; US Patent No. 5,976,782), and the treatment of endothelial cells with atherogenic and angiogenic factors including growth factors and cytokines (e.g., IL-1β, PDGF, TNFα, VEGF), homocysteine, and LDL. *In vitro* angiogenesis models are described in, for example, Black, AF *et al.* (1999) *Cell Biol.Toxicol.* 15:81-90.

Animal based models of cellular proliferative diseases, *e.g.*, tumorigenesis, are well known in the art, and include, for example, non-recombinant and engineered transgenic animals. Models for studying tumorigenesis *in vivo* include carcinogen-induced tumors, injection and/or transplantation of tumor cells into an animal, as well as animals bearing mutations in growth regulatory genes. Models for studying angiogenesis *in vivo* include tumor cell-induced angiogenesis and tumor metastasis (Hoffman, RM (1998-99) *Cancer Metastasis Rev.* 17:271-277; Holash, J *et al.* (1999) *Oncogene* 18:5356-5362; Li, CY *et al.* (2000) *J. Natl Cancer Inst.* 92:143-147), matrix induced angiogenesis (US Patent No. 5,382,514), the disc angiogenesis system (Kowalski, J. *et al.* (1992) *Exp. Mol. Pathol.* 56:1-19), the rodent mesenteric-window angiogenesis assay (Norrby, K (1992) *EXS* 61:282-286), experimental choroidal neovascularization in the rat (Shen, WY *et al.* (1998) *Br. J. Ophthalmol.* 82:1063-1071), and the chick embryo development (Brooks, PC *et al. Methods Mol. Biol.* (1999) 129:257-269) and chick embryo chorioallantoic membrane (CAM) models (McNatt LG *et al.* (1999) *J. Ocul. Pharmacol. Ther.* 15:413-423; Ribatti, D *et al.* (1996) *Int. J. Dev. Biol.* 40:1189-1197), and are reviewed in Ribatti, D and Vacca, A (1999) *Int. J. Biol. Markers* 14:207-213.

This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model. For example, an agent identified as described herein (*e.g.*, a VEGF-G modulating agent, an antisense VEGF-G nucleic acid molecule, a VEGF-G-specific antibody, or a VEGF-G-binding partner) can be used in an animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

B. Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

1. Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the VEGF-G nucleotide sequences, described herein, can be used to map the location of the VEGF-G genes on a chromosome. The mapping of the VEGF-G sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease. The VEGF-G gene has been mapped to human chromosome 11.

Briefly, VEGF-G genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the VEGF-G nucleotide sequences. Computer analysis of the VEGF-G sequences can be used to predict primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the VEGF-G sequences will yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different mammals (*e.g.*, human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but human cells can, the one human chromosome that contains the gene encoding the needed enzyme, will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single

human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes. (D'Eustachio P. *et al.* (1983) *Science* 220:919-924). Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the VEGF-G nucleotide sequences to design oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes. Other mapping strategies which can similarly be used to map a VEGF-G sequence to its chromosome include *in situ* hybridization (described in Fan, Y. *et al.* (1990) *Proc. Natl. Acad. Sci. USA*, 87:6223-27), pre-screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical such as colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases will suffice to get good results at a reasonable amount of time. For a review of this technique, see Verma *et al.*, Human Chromosomes: A Manual of Basic Techniques (Pergamon Press, New York 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding

sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between a gene and a disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, for example, Egeland, J. *et al.* (1987) *Nature*, 325:783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the VEGF-G gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

2. Tissue Typing

The VEGF-G sequences of the present invention can also be used to identify individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent 5,272,057).

Furthermore, the sequences of the present invention can be used to provide an alternative technique which determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the VEGF-G nucleotide sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the present invention can be used to obtain such identification sequences from individuals and from tissue. The VEGF-G nucleotide sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NO:1 or 10 can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO:3 or 12 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

If a panel of reagents from VEGF-G nucleotide sequences described herein is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

3. Use of Partial VEGF-G Sequences in Forensic Biology

DNA-based identification techniques can also be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, for example, a perpetrator of a crime. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, *e.g.*, hair or skin, or body fluids, *e.g.*, blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents, *e.g.*, PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (*i.e.* another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions of SEQ ID NO:1 or 10 are particularly appropriate for this use as greater numbers of polymorphisms occur in the noncoding regions, making it easier to differentiate individuals using this technique. Examples of polynucleotide reagents include the VEGF-G nucleotide sequences or portions thereof, *e.g.*, fragments derived from the noncoding regions of SEQ ID NO:1 or 10 having a length of at least 20 bases, preferably at least 30 bases.

The VEGF-G nucleotide sequences described herein can further be used to provide polynucleotide reagents, *e.g.*, labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue, *e.g.*, a tissue containing endothelial cells. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such VEGF-G probes can be used to identify tissue by species and/or by organ type.

In a similar fashion, these reagents, *e.g.*, VEGF-G primers or probes can be used to screen tissue culture for contamination (*i.e.* screen for the presence of a mixture of different types of cells in a culture).

C. Predictive Medicine:

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining VEGF-G protein and/or nucleic acid expression as well as VEGF-G activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant or unwanted VEGF-G expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with VEGF-G protein, nucleic acid expression or activity. For example, mutations in a VEGF-G gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with VEGF-G protein, nucleic acid expression or activity.

Another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of VEGF-G in clinical trials.

These and other agents are described in further detail in the following sections.

1. Diagnostic Assays

An exemplary method for detecting the presence or absence of VEGF-G protein or nucleic acid in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting VEGF-G protein or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes VEGF-G protein such that the presence of VEGF-G protein or nucleic acid is detected in the biological sample. A preferred agent for detecting VEGF-G mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to VEGF-G mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length VEGF-G nucleic acid, such as the nucleic acid of SEQ ID NO:1, 3, 10 or 12, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length

and sufficient to specifically hybridize under stringent conditions to VEGF-G mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

A preferred agent for detecting VEGF-G protein is an antibody capable of binding to VEGF-G protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or $F(ab')_2$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect VEGF-G mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of VEGF-G mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of VEGF-G protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. *In vitro* techniques for detection of VEGF-G genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of VEGF-G protein include introducing into a subject a labeled anti-VEGF-G antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a serum sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting VEGF-G protein, mRNA, or genomic DNA, such that the presence of VEGF-G protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of VEGF-G protein, mRNA or genomic DNA in the control sample with the presence of VEGF-G protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of VEGF-G in a biological sample. For example, the kit can comprise a labeled compound or agent capable of detecting VEGF-G protein or mRNA in a biological sample; means for determining the amount of VEGF-G in the sample; and means for comparing the amount of VEGF-G in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect VEGF-G protein or nucleic acid.

2. Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant or unwanted VEGF-G expression or activity. As used herein, the term "aberrant" includes a VEGF-G expression or activity which deviates from the wild type VEGF-G expression or activity. Aberrant expression or activity includes increased or decreased expression or activity, as well as expression or activity which does not follow the wild type developmental pattern of expression or the subcellular pattern of expression. For example, aberrant VEGF-G expression or activity is intended to include the cases in which a mutation in the VEGF-G gene causes the VEGF-G gene to be under-expressed or over-expressed and situations in which such mutations result in a non-functional VEGF-G protein or a protein which does not function in a wild-type fashion, *e.g.*, a protein which does not interact with a VEGF-G substrate, *e.g.*, a VEGF receptor, or one which interacts with a non-VEGF-G substrate. As used herein, the term "unwanted" includes an unwanted phenomenon involved in a biological response such

as pain or deregulated cell proliferation. For example, the term unwanted includes a VEGF-G expression or activity which is undesirable in a subject.

The assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with a misregulation in VEGF-G protein activity or nucleic acid expression, such as a cell proliferation and/or differentiation disorder. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disorder associated with a misregulation in VEGF-G protein activity or nucleic acid expression, such as a cell proliferation and/or differentiation disorder. Thus, the present invention provides a method for identifying a disease or disorder associated with aberrant or unwanted VEGF-G expression or activity in which a test sample is obtained from a subject and VEGF-G protein or nucleic acid (*e.g.*, mRNA or genomic DNA) is detected, wherein the presence of VEGF-G protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant or unwanted VEGF-G expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant or unwanted VEGF-G expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a cell proliferation and/or differentiation disorder. Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant or unwanted VEGF-G expression or activity in which a test sample is obtained and VEGF-G protein or nucleic acid expression or activity is detected (*e.g.*, wherein the abundance of VEGF-G protein or nucleic acid expression or activity is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant or unwanted VEGF-G expression or activity).

The methods of the invention can also be used to detect genetic alterations in a VEGF-G gene, thereby determining if a subject with the altered gene is at risk for a disorder characterized by misregulation in VEGF-G protein activity or nucleic acid expression, such as a cell proliferation and/or differentiation disorder. In preferred embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic alteration characterized by at least one of an alteration affecting the integrity of a gene encoding a VEGF-G-protein, or the mis-expression of the VEGF-G gene. For example, such genetic alterations can be detected by ascertaining the existence of at least one of 1) a deletion of one or more nucleotides from a VEGF-G gene; 2) an addition of one or more nucleotides to a VEGF-G gene; 3) a substitution of one or more nucleotides of a VEGF-G gene, 4) a chromosomal rearrangement of a VEGF-G gene; 5) an alteration in the level of a messenger RNA transcript of a VEGF-G gene, 6) aberrant modification of a VEGF-G gene, such as of the methylation pattern of the genomic DNA, 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a VEGF-G gene, 8) a non-wild type level of a VEGF-G-protein, 9) allelic loss of a VEGF-G gene, and 10) inappropriate post-translational modification of a VEGF-G-protein. As described herein, there are a large number of assays known in the art which can be used for detecting alterations in a VEGF-G gene. A preferred biological sample is a tissue or serum sample isolated by conventional means from a subject.

In certain embodiments, detection of the alteration involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, *e.g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, *e.g.*, Landegran *et al.* (1988) *Science* 241:1077-1080; and Nakazawa *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:360-364), the latter of which can be particularly useful for detecting point mutations in the VEGF-G-gene (see Abravaya *et al.* (1995) *Nucleic Acids Res* .23:675-682). This method can include the steps of collecting a sample of cells from a subject, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a VEGF-G gene under conditions such that hybridization and amplification of the VEGF-G-gene (if present) occurs, and

detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli, J.C. *et al.*, (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh, D.Y. *et al.*, (1989) *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi, P.M. *et al.* (1988) *Bio-Technology* 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a VEGF-G gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, for example, U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in VEGF-G can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotides probes (Cronin, M.T. *et al.* (1996) *Human Mutation* 7: 244-255; Kozal, M.J. *et al.* (1996) *Nature Medicine* 2: 753-759). For example, genetic mutations in VEGF-G can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, M.T. *et al. supra.* Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array

that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the VEGF-G gene and detect mutations by comparing the sequence of the sample VEGF-G with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxam and Gilbert ((1977) *Proc. Natl. Acad. Sci. USA* 74:560) or Sanger ((1977) *Proc. Natl. Acad. Sci. USA* 74:5463). It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) *Biotechniques* 19:448), including sequencing by mass spectrometry (see, *e.g.*, PCT International Publication No. WO 94/16101; Cohen *et al.* (1996) *Adv. Chromatogr.* 36:127-162; and Griffin *et al.* (1993) *Appl. Biochem. Biotechnol.* 38:147-159).

Other methods for detecting mutations in the VEGF-G gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers *et al.* (1985) *Science* 230:1242). In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type VEGF-G sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent which cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, for example, Cotton *et al.* (1988) *Proc. Natl Acad Sci USA* 85:4397; Saleeba *et al.*

(1992) *Methods Enzymol.* 217:286-295. In a preferred embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in VEGF-G cDNAs obtained from samples of cells. For example, the mutY enzyme of E. coli cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu *et al.* (1994) *Carcinogenesis* 15:1657-1662). According to an exemplary embodiment, a probe based on a VEGF-G sequence, *e.g.*, a wild-type VEGF-G sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. See, for example, U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in VEGF-G genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (orita *et al.* (1989) *Proc Natl. Acad. Sci USA*: 86:2766, see also Cotton (1993) *Mutat. Res.* 285:125-144; and Hayashi (1992) *Genet. Anal. Tech. Appl.* 9:73-79). Single-stranded DNA fragments of sample and control VEGF-G nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen *et al.* (1991) *Trends Genet* 7:5).

In yet another embodiment the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers *et al.* (1985) *Nature* 313:495). When

DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) *Biophys Chem* 265:12753).

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki *et al.* (1986) *Nature* 324:163); Saiki *et al.* (1989) *Proc. Natl Acad. Sci USA* 86:6230). Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs *et al.* (1989) *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini *et al.* (1992) *Mol. Cell Probes* 6:1). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany (1991) *Proc. Natl. Acad. Sci USA* 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose

patients exhibiting symptoms or family history of a disease or illness involving a VEGF-G gene.

Furthermore, any cell type or tissue in which VEGF-G is expressed may be utilized in the prognostic assays described herein.

3. Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (*e.g.*, drugs) on the expression or activity of a VEGF-G protein (*e.g.*, the modulation of cell growth, differentiation, migration, and/or apoptosis mechanisms) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase VEGF-G gene expression, protein levels, or upregulate VEGF-G activity, can be monitored in clinical trials of subjects exhibiting decreased VEGF-G gene expression, protein levels, or downregulated VEGF-G activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease VEGF-G gene expression, protein levels, or downregulate VEGF-G activity, can be monitored in clinical trials of subjects exhibiting increased VEGF-G gene expression, protein levels, or upregulated VEGF-G activity. In such clinical trials, the expression or activity of a VEGF-G gene, and preferably, other genes that have been implicated in, for example, a VEGF-G-associated disorder can be used as a "read out" or markers of the phenotype of a particular cell.

For example, and not by way of limitation, genes, including VEGF-G, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) which modulates VEGF-G activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on VEGF-G-associated disorders (*e.g.*, disorders characterized by deregulated cell growth, differentiation and/or migration mechanisms), for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of VEGF-G and other genes implicated in the VEGF-G-associated disorder, respectively. The levels of gene expression (*e.g.*, a gene expression pattern) can be quantified by northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of

activity of VEGF-G or other genes. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during treatment of the individual with the agent.

In a preferred embodiment, the present invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) including the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of a VEGF-G protein, mRNA, or genomic DNA in the preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the VEGF-G protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the VEGF-G protein, mRNA, or genomic DNA in the pre-administration sample with the VEGF-G protein, mRNA, or genomic DNA in the post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of VEGF-G to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of VEGF-G to lower levels than detected, *i.e.* to decrease the effectiveness of the agent. According to such an embodiment, VEGF-G expression or activity may be used as an indicator of the effectiveness of an agent, even in the absence of an observable phenotypic response.

D. Methods of Treatment:

The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant or unwanted VEGF-G expression or activity. With regards to both prophylactic and therapeutic methods of treatment, such treatments may be specifically tailored or modified, based on knowledge obtained from the field of

pharmacogenomics. "Pharmacogenomics", as used herein, refers to the application of genomics technologies such as gene sequencing, statistical genetics, and gene expression analysis to drugs in clinical development and on the market. More specifically, the term refers the study of how a patient's genes determine his or her response to a drug (*e.g.*, a patient's "drug response phenotype", or "drug response genotype".) Thus, another aspect of the invention provides methods for tailoring an individual's prophylactic or therapeutic treatment with either the VEGF-G molecules of the present invention or VEGF-G modulators according to that individual's drug response genotype. Pharmacogenomics allows a clinician or physician to target prophylactic or therapeutic treatments to patients who will most benefit from the treatment and to avoid treatment of patients who will experience toxic drug-related side effects.

1. Prophylactic Methods

In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with an aberrant or unwanted VEGF-G expression or activity, by administering to the subject a VEGF-G or an agent which modulates VEGF-G expression or at least one VEGF-G activity. Subjects at risk for a disease which is caused or contributed to by aberrant or unwanted VEGF-G expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the VEGF-G aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of VEGF-G aberrancy, for example, a VEGF-G, VEGF-G agonist or VEGF-G antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein.

2. Therapeutic Methods

Another aspect of the invention pertains to methods of modulating VEGF-G expression or activity for therapeutic purposes. Accordingly, in an exemplary embodiment, the modulatory method of the invention involves contacting a cell with a VEGF-G or agent that modulates one or more of the activities of VEGF-G protein

activity associated with the cell. An agent that modulates VEGF-G protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring target molecule of a VEGF-G protein (*e.g.*, a VEGF-G substrate or receptor), a VEGF-G antibody, a VEGF-G agonist or antagonist, a peptidomimetic of a VEGF-G agonist or antagonist, or other small molecule. In one embodiment, the agent stimulates one or more VEGF-G activities. Examples of such stimulatory agents include active VEGF-G protein and a nucleic acid molecule encoding VEGF-G that has been introduced into the cell. In another embodiment, the agent inhibits one or more VEGF-G activities. Examples of such inhibitory agents include antisense VEGF-G nucleic acid molecules, anti-VEGF-G antibodies, and VEGF-G inhibitors. These modulatory methods can be performed *in vitro* (*e.g.*, by culturing the cell with the agent) or, alternatively, *in vivo* (*e.g.*, by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant or unwanted expression or activity of a VEGF-G protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (*e.g.*, an agent identified by a screening assay described herein), or combination of agents that modulates (*e.g.*, upregulates or downregulates) VEGF-G expression or activity. In another embodiment, the method involves administering a VEGF-G protein or nucleic acid molecule as therapy to compensate for reduced, aberrant, or unwanted VEGF-G expression or activity.

Stimulation of VEGF-G activity is desirable in situations in which VEGF-G is abnormally downregulated and/or in which increased VEGF-G activity is likely to have a beneficial effect. For example, stimulation of VEGF-G activity is desirable in situations in which a VEGF-G is downregulated and/or in which increased VEGF-G activity is likely to have a beneficial effect. Likewise, inhibition of VEGF-G activity is desirable in situations in which VEGF-G is abnormally upregulated and/or in which decreased VEGF-G activity is likely to have a beneficial effect.

3. Pharmacogenomics

The VEGF-G molecules of the present invention, as well as agents, or modulators which have a stimulatory or inhibitory effect on VEGF-G activity (*e.g.*,

VEGF-G gene expression) as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) VEGF-G-associated disorders (*e.g.*, cell proliferation and/or differentiation disorders, or disorders characterized by aberrant angiogenesis) associated with aberrant or unwanted VEGF-G activity. In conjunction with such treatment, pharmacogenomics (*i.e.*, the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, a physician or clinician may consider applying knowledge obtained in relevant pharmacogenomics studies in determining whether to administer a VEGF-G molecule or VEGF-G modulator as well as tailoring the dosage and/or therapeutic regimen of treatment with a VEGF-G molecule or VEGF-G modulator.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, for example, Eichelbaum, M. *et al.* (1996) *Clin. Exp. Pharmacol. Physiol.* 23(10-11) :983-985 and Linder, M.W. *et al.* (1997) *Clin. Chem.* 43(2):254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare genetic defects or as naturally-occurring polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

One pharmacogenomics approach to identifying genes that predict drug response, known as "a genome-wide association", relies primarily on a high-resolution map of the human genome consisting of already known gene-related markers (*e.g.*, a "bi-allelic" gene marker map which consists of 60,000-100,000 polymorphic or variable sites on the human genome, each of which has two variants.) Such a high-resolution

genetic map can be compared to a map of the genome of each of a statistically significant number of patients taking part in a Phase II/III drug trial to identify markers associated with a particular observed drug response or side effect. Alternatively, such a high resolution map can be generated from a combination of some ten-million known single nucleotide polymorphisms (SNPs) in the human genome. As used herein, a "SNP" is a common alteration that occurs in a single nucleotide base in a stretch of DNA. For example, a SNP may occur once per every 1000 bases of DNA. A SNP may be involved in a disease process, however, the vast majority may not be disease-associated. Given a genetic map based on the occurrence of such SNPs, individuals can be grouped into genetic categories depending on a particular pattern of SNPs in their individual genome. In such a manner, treatment regimens can be tailored to groups of genetically similar individuals, taking into account traits that may be common among such genetically similar individuals.

Alternatively, a method termed the "candidate gene approach", can be utilized to identify genes that predict drug response. According to this method, if a gene that encodes a drug's target is known (*e.g.*, a VEGF-G protein of the present invention), all common variants of that gene can be fairly easily identified in the population and it can be determined if having one version of the gene versus another is associated with a particular drug response.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses.

If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Alternatively, a method termed the "gene expression profiling", can be utilized to identify genes that predict drug response. For example, the gene expression of an animal dosed with a drug (*e.g.*, a VEGF-G molecule or VEGF-G modulator of the present invention) can give an indication whether gene pathways related to toxicity have been turned on.

Information generated from more than one of the above pharmacogenomics approaches can be used to determine appropriate dosage and treatment regimens for prophylactic or therapeutic treatment an individual. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a VEGF-G molecule or VEGF-G modulator, such as a modulator identified by one of the exemplary screening assays described herein.

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of the Sequence Listing, figures, and all references, patents and published patent applications cited throughout this application are incorporated herein by reference.

EXAMPLES

EXAMPLE 1: IDENTIFICATION AND CHARACTERIZATION OF HUMAN AND MOUSE VEGF-G cDNA

In this example, the identification and characterization of the gene encoding human (clone jthbb079f06) and mouse VEGF-G is described.

Isolation of the human and mouse VEGF-G cDNA

The invention is based, at least in part, on the discovery of genes encoding novel members of the VEGF family. The human VEGF-G cDNA was isolated from a human osteoblast library.

The nucleotide sequence encoding the human VEGF-G protein (clone jthbb079f06) is shown in Figure 1 and is set forth as SEQ ID NO:1. The full length protein encoded by this nucleic acid comprises about 370 amino acids and has the amino acid sequence shown in Figure 1 and set forth as SEQ ID NO:2. The coding region (open reading frame) of SEQ ID NO:1 is set forth as SEQ ID NO:3.

The nucleotide sequence encoding the mouse VEGF-G protein is shown in Figure 8 and is set forth as SEQ ID NO:10. The full length protein encoded by this nucleic acid comprises about 370 amino acids and has the amino acid sequence shown in Figure 8 and set forth as SEQ ID NO:11. The coding region (open reading frame) of SEQ ID NO:10 is set forth as SEQ ID NO:12.

Analysis of the VEGF-G Molecules

A search was performed against the HMM database resulting in the identification of a CUB domain in the amino acid sequence of human VEGF-G (SEQ ID NO:2) at about residues 53-167 of SEQ ID NO:2. The results of the search are set forth in Figure 2. A CUB domain was also identified in the amino acid sequence of mouse VEGF-G (SEQ ID NO:11) at about residues 53-167 of SEQ ID NO:11. The results of the search are set forth in Figure 9.

The human VEGF-G protein is predicted to have a signal peptide from amino acid residues 1-19 of SEQ ID NO:2. Accordingly, a mature human VEGF-G protein is predicted to include amino acid residues 20-370 of SEQ ID NO:2. The mouse VEGF-G protein is predicted to have a signal peptide from amino acid residues 1-23 of SEQ ID NO:11. Accordingly, a mature mouse VEGF-G protein is predicted to include amino acid residues 24-370 of SEQ ID NO:11.

The VEGF-G protein is also predicted to have at least one N-glycosylation site, at about amino acid residues 276-279 of SEQ ID NO:2, and at about amino acid residues 14-17 and 276-279 of SEQ ID NO:11. In addition, the VEGF-G protein is predicted to

have at least one N-myristoylation site, at about amino acid residues 100-105, 192-197, and 303-308 of SEQ ID NO:2, and at about amino acid residues 100-105 and 303-308 of SEQ ID NO:11.

The VEGF-G protein is predicted to have at least one cAMP and cGMP dependent protein kinase phosphorylation site, at about amino acid residues 268-271 of SEQ ID NO:2, and at about amino acid residues 268-271 of SEQ ID NO:11.

The VEGF-G protein is predicted to have at least one protein kinase C phosphorylation site, at about amino acid residues 17-19, 29-31, 66-68, 80-82, 150-152, 243-245, 273-275, 320-322, 323-325, and 365-367 of SEQ ID NO:2, and at about amino acid residues 29-31, 66-68, 141-143, 150-152, 273-275, 320-322, 323-325, and 365-367 of SEQ ID NO:11.

The VEGF-G protein is predicted to have at least one casein kinase II phosphorylation site, at about amino acid residues 17-20, 168-171, 181-184, 199-202, 219-222, 231-234, 250-253, and 256-259 of SEQ ID NO:2, and at about amino acid residues 168-171, 181-184, 199-202, 213-216, 219-222, 231-234, 250-253, and 256-259 of SEQ ID NO:11.

The VEGF-G protein is predicted to have at least one tyrosine kinase phosphorylation site, at about amino acid residues 262-270 of SEQ ID NO:2, and at about amino acid residues 262-270 of SEQ ID NO:11.

The open reading frame of the human VEGF-G gene was globally aligned with the open reading frame of the mouse VEGF-G gene using the GAP program in the GCG software package, using a nwsgapdna.cmp matrix and a gap weight of 12 and a length weight of 4. The results showed a 85.586% identity between the two sequences (see Figure 10). In addition, the nucleotide sequence of the human VEGF-G gene was locally aligned with the nucleotide sequence of the mouse VEGF-G gene using the GAP program in the GCG software package, using a nwsgapdna.cmp matrix and a gap weight of 12 and a length weight of 4. The results showed a 74.592% identity between the two sequences.

The human VEGF-G protein was globally aligned with the mouse VEGF-G protein using the GAP program in the GCG software package, using a Blosum 62 matrix and a gap weight of 12 and a length weight of 4. The results showed a 84.865% identity between the two sequences (see Figure 11).

Tissue Distribution of VEGF-G mRNA

Northern blot hybridizations with various human tissue RNA samples revealed that VEGF-G mRNA was strongly expressed in the ovary, heart, pancreas, and fetal kidney, moderately expressed in the prostate, testis, small intestine, spinal cord, trachea, placenta and kidney, and weakly expressed in the colon, brain, skeletal muscle, and liver. Within the endocrine system, VEGF-G mRNA was strongly expressed in the adrenal medulla and the adrenal cortex. In addition, VEGF-G mRNA was weakly expressed in various immune system tissues such as spleen, lymph node, thymus, peripheral blood leukocytes, bone marrow, and fetal liver. VEGF-G mRNA was also expressed in the thyroid and stomach, but was not expressed in adult or fetal lung, or fetal brain. Northern blot analysis of mouse tissues revealed strong expression of VEGF-G mRNA in the heart and kidney, moderate expression in the brain, and no expression in the spleen, lung, liver, skeletal muscle and testis.

Tissue Distribution of VEGF-G cDNA

VEGF-G expression was also determined by PCR analysis of cDNA libraries from various tissues and cell lines. Detecting expression by a library array procedure entailed preparing a PCR mixture including Taq Polymerase, dNTPs, and PCR buffer, and adding a vector primer, a primer internal to the gene of interest, and an aliquot of a library in which expression was to be tested. This procedure was performed with many libraries at a time in a 96 well PCR tray, with 80 or more wells containing libraries and a control well in which the above primers were combined with the clone of interest itself. The control well served as an indicator of the fragment size to be expected in the library wells, in the event the clone of interest was expressed within. Amplification was performed in a PCR machine, employing standard PCR conditions for denaturing, annealing, and elongation, and the resultant mixture was mixed with an appropriate

loading dye and run on an ethidium bromide-stained agarose gel. The gel was later viewed with UV light after the DNA loaded within its lanes had time to migrate into the gels. Lanes in which a band corresponding with the control band was visible indicated the libraries in which the clone of interest was expressed.

VEGF-G was expressed in choroid plexus (MCPdL) and lung Bleomycin model day 7, and weakly expressed in long term bone marrow cells, MLTC-1 (mouse Leydig tumor cells), lung, brain, and lung Gonzolo day 15, 3 hours.

EXAMPLE 2 REGULATION OF VEGF-G EXPRESSION IN HUMAN ENDOTHELIAL CELLS

The expression of VEGF-G in human endothelial cells was analyzed by TaqMan® Quantitative Polymerase Chain Reaction.

Probes were designed by PrimerExpress software (PE Biosystems) based on the sequence of the VEGF-G gene. Each VEGF-G gene probe was labeled using FAM (6-carboxyfluorescein), and the β2-microglobulin reference probe was labeled with a different fluorescent dye, VIC. The differential labeling of the target gene and internal reference gene thus enabled measurement in same well. Forward and reverse primers and the probes for both β2-microglobulin and target gene were added to the TaqMan® Universal PCR Master Mix (PE Applied Biosystems). Although the final concentration of primer and probe could vary, each was internally consistent within a given experiment. A typical experiment contained 200nM of forward and reverse primers plus 100nM probe for β-2 microglobulin and 600 nM forward and reverse primers plus 200 nM probe for the target gene. TaqMan matrix experiments were carried out on an ABI PRISM 7700 Sequence Detection System (PE Applied Biosystems). The thermal cycler conditions were as follows: hold for 2 minute at 50°C and 10 minute at 95°C, followed by two-step PCR for 40 cycles of 95°C for 15 seconds followed by 60°C for 1 minute.

A comparative Ct method is used for the relative quantitation of gene expression. The following method was used to quantitatively calculate VEGF-G gene expression in the various samples relative to β-2 microglobulin expression in the same sample. The threshold cycle (Ct) value is defined as the cycle at which a statistically significant

increase in fluorescence is detected. A lower Ct value is indicative of a higher mRNA concentration. The Ct value of the VEGF-G gene is normalized by subtracting the Ct value of the β-2 microglobulin gene to obtain a ΔCt value using the following formula:

$$\Delta Ct = Ct_{VEGF\text{-}G} - Ct_{\beta\text{-}2\ microglobulin}$$

Expression is then calibrated against a cDNA sample showing a comparatively low level of expression of the VEGF-G gene. The ΔCt value for the calibrator sample is then subtracted from ΔCt for each tissue sample according to the following formula:

$$\Delta\Delta Ct = \Delta Ct_{sample} - \Delta Ct_{calibrator}$$

Relative expression is then calculated using the arithmetic formula given by $2^{-\Delta\Delta Ct}$.

Human umbilical vein endothelial cells (HUVECs) were cultured *in vitro* under standard conditions. Experimental cultures were then exposed to either laminar shear stress (LSS) conditions or interleukin (IL)-1.

Cultured HUVEC monolayers were exposed to laminar sheer stress by culturing the cells in a specialized apparatus containing liquid culture medium. Static cultures grown in the same medium served as controls. The *in vitro* LSS treatment at 10 dyns/cm^2 was designed to simulate the shear stress generated by blood flow in a straight, healthy artery such as the internal mammary artery. Alternatively, HUVEC cultures were treated with human IL-1β, a factor known to be involved in the inflammatory response, in order to mimic the physiologic conditions involved in the atherosclerotic state. Stimulation of endothelial cells with IL-1 induces the expression of several inflammatory markers. Experimental and control cells were harvested and analyzed for gene expression at 1, 6 and 24 hours.

Exposure of HUVECs to LSS resulted in a reduction in VEGF-G gene expression at 1 and 6 hours, and an induction in VEGF-G levels following 24 hours of LSS treatment, as compared to static controls. In addition, IL-1 stimulation of HUVECs resulted in a time dependent decrease in VEGF-G expression.

In another study, primary cultures of human microvascular endothelial cells (HMVEC) were plated on Matrigel to induce tube formation and incubated for 5 or 25 hours, or were grown on plastic tissue culture dishes as a control. Control cells include sub-confluent HMVEC that were grown on tissue culture dishes in 5% serum and

incubated in the presence of growth factors for 25 hours, and confluent HMVEC that were grown on tissue culture dishes in 5% serum and incubated in the absence of growth factors for 25 hours. VEGF-G gene expression was up-regulated after 25 hours of incubation on Matrigel, at which time the cells are performing angiogenesis-related functions, *e.g.*, cell migration and tube formation.

Collectively, these data indicate that VEGF-G may be involved in the regulation of endothelial cell processes such as growth, proliferation, differentiation, migration and tube formation.

EXAMPLE 3: VEGF-G IS A SECRETED PROTEIN

This example describes a secretion assay that was performed to determine if VEGF-G is secreted from cells. Briefly, the secretion assay was performed as follows: $8x10^5$ 293T cells were plated per well in a 6-well plate and the cells were incubated in growth medium (DMEM, 10% fetal bovine serum, penicillin/strepomycin) at 37°C, 5% CO_2 overnight. 293T cells were transfected with 2 µg per well of full-length VEGF-G in the pMET7 vector and 10 µg per well of LipofectAMINE (GIBCO/BRL Cat. # 18324-012) according to the manufacturer's instructions. After 5 hours, the transfection media was removed and fresh growth medium was added to allow the cells to recover overnight. The medium was removed and each well was gently washed twice with DMEM without methionine and cysteine (ICN Cat. # 16-424-54). Subsequently, 1 ml DMEM without methionine and cysteine supplemented with 50 µCi Trans-^{35}S-Label (^{35}S-L-methionine/^{35}S-L-cysteine; ICN Cat. # 51006) was added to each well and the cells were incubated at 37°C, 5% CO_2. A 150 µl aliquot of conditioned medium was obtained and 150 µl of 2X SDS sample buffer was added to the aliquot. The sample was heat-inactivated and loaded on a 4-20% SDS-PAGE gel. The gel was fixed and the presence of secreted protein was detected by autoradiography. The secretion assay revealed that VEGF-G is a secreted protein.

EXAMPLE 4: VEGF-G BINDING TO ENDOTHELIAL CELLS

The binding of VEGF-G to endothelial cells was assessed using an alkaline phosphatase-VEGF-G fusion protein (AP-VEGF-G). *In vitro* studies of AP-VEGF-G binding to HDMEC (human dermal microvascular endothelial cells), bACE cells (bovine adrenal cortical capillary endothelial cells) and BAE (bovine aortic endothelial cells) were performed with Phospha-Light chemilluminescent assay system (Tropix, Inc. Bedford, MA). bACE cells were plated into gelatinized 96-well plates (3000 cells/well) and allowed to grow to confluence. The cells were then fixed with acetone. AP-VEGF-G was incubated with the cells for 1 hour, and specific binding was detected with a microplate luminometer according to the manufacturer's instructions.

The results from these experiments indicated selective, high affinity binding of VEGF-G to endothelial cells in culture. When human smooth muscle cells, human prostate carcinoma cells (PC-3), human colon carcinoma cells (LS-174T) and human melanoma cells (SK-MEL-5) were examined, only background binding was detected. These data suggest that VEGF-G binds to endothelial cells with high affinity, and may thus selectively exert its effects on endothelial cells.

EXAMPLE 5: EXPRESSION OF RECOMBINANT VEGF-G PROTEIN IN BACTERIAL CELLS

In this example, VEGF-G is expressed as a recombinant glutathione-S-transferase (GST) fusion polypeptide in *E. coli* and the fusion polypeptide is isolated and characterized. Specifically, VEGF-G is fused to GST and this fusion polypeptide is expressed in *E. coli*, *e.g.*, strain PEB199. Expression of the GST-VEGF-G fusion protein in PEB199 is induced with IPTG. The recombinant fusion polypeptide is purified from crude bacterial lysates of the induced PEB199 strain by affinity chromatography on glutathione beads. Using polyacrylamide gel electrophoretic analysis of the polypeptide purified from the bacterial lysates, the molecular weight of the resultant fusion polypeptide is determined.

EXAMPLE 6: EXPRESSION OF RECOMBINANT VEGF-G PROTEIN IN COS CELLS

To express the VEGF-G gene in COS cells, the pcDNA/Amp vector by Invitrogen Corporation (San Diego, CA) is used. This vector contains an SV40 origin of replication, an ampicillin resistance gene, an *E. coli* replication origin, a CMV promoter followed by a polylinker region, and an SV40 intron and polyadenylation site. A DNA fragment encoding the entire VEGF-G protein and an HA tag (Wilson *et al.* (1984) *Cell* 37:767) or a FLAG tag fused in-frame to its 3' end of the fragment is cloned into the polylinker region of the vector, thereby placing the expression of the recombinant protein under the control of the CMV promoter.

To construct the plasmid, the VEGF-G DNA sequence is amplified by PCR using two primers. The 5' primer contains the restriction site of interest followed by approximately twenty nucleotides of the VEGF-G coding sequence starting from the initiation codon; the 3' end sequence contains complementary sequences to the other restriction site of interest, a translation stop codon, the HA tag or FLAG tag and the last 20 nucleotides of the VEGF-G coding sequence. The PCR amplified fragment and the pCDNA/Amp vector are digested with the appropriate restriction enzymes and the vector is dephosphorylated using the CIAP enzyme (New England Biolabs, Beverly, MA). Preferably the two restriction sites chosen are different so that the VEGF-G gene is inserted in the correct orientation. The ligation mixture is transformed into *E. coli* cells (strains HB101, DH5α, SURE, available from Stratagene Cloning Systems, La Jolla, CA, can be used), the transformed culture is plated on ampicillin media plates, and resistant colonies are selected. Plasmid DNA is isolated from transformants and examined by restriction analysis for the presence of the correct fragment.

COS cells are subsequently transfected with the VEGF-G-pcDNA/Amp plasmid DNA using the calcium phosphate or calcium chloride co-precipitation methods, DEAE-dextran-mediated transfection, lipofection, or electroporation. Other suitable methods for transfecting host cells can be found in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual, 2nd. ed., Cold Spring Harbor Laboratory*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989. The expression

of the VEGF-G polypeptide is detected by radiolabelling (^{35}S-methionine or ^{35}S-cysteine available from NEN, Boston, MA, can be used) and immunoprecipitation (Harlow, E. and Lane, D. *Antibodies: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1988) using an HA specific monoclonal antibody. Briefly, the cells are labelled for 8 hours with ^{35}S-methionine (or ^{35}S-cysteine). The culture media are then collected and the cells are lysed using detergents (RIPA buffer, 150 mM NaCl, 1% NP-40, 0.1% SDS, 0.5% DOC, 50 mM Tris, pH 7.5). Both the cell lysate and the culture media are precipitated with an HA specific monoclonal antibody. Precipitated polypeptides are then analyzed by SDS-PAGE.

Alternatively, DNA containing the VEGF-G coding sequence is cloned directly into the polylinker of the pCDNA/Amp vector using the appropriate restriction sites. The resulting plasmid is transfected into COS cells in the manner described above, and the expression of the VEGF-G polypeptide is detected by radiolabelling and immunoprecipitation using a VEGF-G specific monoclonal antibody.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

What is claimed:

1. An isolated nucleic acid molecule selected from the group consisting of:

(a) a nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:1 or 3, or a complement thereof;

(b) a nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:10 or 12, or a complement thereof.

2. An isolated nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 2 or 11, or a complement thereof.

3. An isolated nucleic acid molecule comprising the nucleotide sequence contained in the plasmid deposited with ATCC® as Accession Number _____ .

4. An isolated nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 2 or 11, or a complement thereof.

5. An isolated nucleic acid molecule selected from the group consisting of:

a) a nucleic acid molecule comprising a nucleotide sequence which is at least 60% identical to the nucleotide sequence of SEQ ID NO:1, 3, 10 or 12, or a complement thereof;

b) a nucleic acid molecule comprising a fragment of at least 50 nucleotides of a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1, 3, 10 or 12, or a complement thereof;

c) a nucleic acid molecule which encodes a polypeptide comprising an amino acid sequence at least about 60% identical to the amino acid sequence of SEQ ID NO:2 or 11; and

d) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2 or 11, wherein the

fragment comprises at least 15 contiguous amino acid residues of the amino acid sequence of SEQ ID NO:2 or 11.

6. An isolated nucleic acid molecule which hybridizes to the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 under stringent conditions.

7. An isolated nucleic acid molecule comprising the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5, and a nucleotide sequence encoding a heterologous polypeptide.

8. A vector comprising the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5.

9. The vector of claim 8, which is an expression vector.

10. A host cell transfected with the expression vector of claim 9.

11. A method of producing a polypeptide comprising culturing the host cell of claim 10 in an appropriate culture medium to, thereby, produce the polypeptide.

12. An isolated polypeptide selected from the group consisting of:

a) a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2 or 11, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2 or 11;

b) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2 or 11, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule consisting of SEQ ID NO:1, 3, 10 or 12 under stringent conditions;

c) a polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 60 % identical to a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1, 3, 10 or 12; and

d) a polypeptide comprising an amino acid sequence which is at least 60% identical to the amino acid sequence of SEQ ID NO:2 or 11.

13. The isolated polypeptide of claim 12 comprising the amino acid sequence of SEQ ID NO:2 or 11.

14. The polypeptide of claim 12, further comprising heterologous amino acid sequences.

15. An antibody which selectively binds to a polypeptide of claim 12.

16. A method for detecting the presence of a polypeptide of claim 12 in a sample comprising:

a) contacting the sample with a compound which selectively binds to the polypeptide; and

b) determining whether the compound binds to the polypeptide in the sample to thereby detect the presence of a polypeptide of claim 12 in the sample.

17. The method of claim 16, wherein the compound which binds to the polypeptide is an antibody.

18. A kit comprising a compound which selectively binds to a polypeptide of claim 12 and instructions for use.

19. A method for detecting the presence of a nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 in a sample comprising:

a) contacting the sample with a nucleic acid probe or primer which selectively hybridizes to the nucleic acid molecule; and

b) determining whether the nucleic acid probe or primer binds to a nucleic acid molecule in the sample to thereby detect the presence of a nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 in the sample.

20. The method of claim 19, wherein the sample comprises mRNA molecules and is contacted with a nucleic acid probe.

21. A kit comprising a compound which selectively hybridizes to a nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 and instructions for use.

22. A method for identifying a compound which binds to a polypeptide of claim 12 comprising:

a) contacting the polypeptide, or a cell expressing the polypeptide with a test compound; and

b) determining whether the polypeptide binds to the test compound.

23. The method of claim 22, wherein the binding of the test compound to the polypeptide is detected by a method selected from the group consisting of:

a) detection of binding by direct detection of test compound/polypeptide binding;

b) detection of binding using a competition binding assay; and

c) detection of binding using an assay for VEGF-G activity.

24. A method for modulating the activity of a polypeptide of claim 12 comprising contacting the polypeptide or a cell expressing the polypeptide with a compound which binds to the polypeptide in a sufficient concentration to modulate the activity of the polypeptide.

25. A method for identifying a compound which modulates the activity of a polypeptide of claim 12 comprising:

a) contacting a polypeptide of claim 12 with a test compound; and

b) determining the effect of the test compound on the activity of the polypeptide to thereby identify a compound which modulates the activity of the polypeptide.

```
GTCGACCCACGCGTCCGGGAAGTGCGGCACCCACACGCGCTCGGAAAGTTCAGCATGCAGGAAGTTTGGGGAGAGCTCG   79

GCGATTAGCACAGCGACCCGGGCCAGCGCAGGGCGAGCGCAGGCGGCGAGAGCGCAGGGCGGCGCGGCGTCGGTCCCGG  158

                                                          M   H   R   L   I   F      6
GAGCAGAACCCGGCTTTTTCTTGGAGCGACGCTGTCTCTAGTCGCTGATCCCAA ATG CAC CGG CTC ATC TTT    230

 V   Y   T   L   I   C   A   N   F   C   S   C   R   D   T   S   A   T   P   Q     26
GTC TAC ACT CTA ATC TGC GCA AAC TTT TGC AGC TGT CGG GAC ACT TCT GCA ACC CCG CAG    290

 S   A   S   I   K   A   L   R   N   A   N   L   R   R   D   E   S   N   H   L     46
AGC GCA TCC ATC AAA GCT TTG CGC AAC GCC AAC CTC AGG CGA CAT GAG AGC AAT CAC CTC    350

 T   D   L   Y   R   R   D   E   T   I   Q   V   K   G   N   G   Y   V   Q   S     66
ACA GAC TTG TAC CGA AGA GAT GAG ACC ATC CAG GTG AAA GGA AAC GGC TAC GTG CAG AGT    410

 P   R   F   P   N   S   Y   P   R   N   L   L   L   T   W   R   L   H   S   Q     86
CCT AGA TTC CCG AAC AGC TAC CCC AGG AAC CTG CTC CTG ACA TGG CGG CTT CAC TCT CAG    470

 E   N   T   R   I   Q   L   V   F   D   N   Q   F   G   L   E   E   A   E   N    106
GAG AAT ACA CGG ATA CAG CTA GTG TTT GAC AAT CAG TTT GGA TTA GAG GAA GCA GAA AAT    530

 D   I   C   R   Y   D   F   V   E   V   E   D   I   S   E   T   S   T   I   I    126
GAT ATC TGT AGG TAT GAT TTT GTG GAA GTT GAA GAT ATA TCC GAA ACC AGT ACC ATT ATT    590

 R   G   R   W   C   G   H   K   E   V   P   P   R   I   K   S   R   T   N   Q    146
AGA GGA CGA TGG TGT GGA CAC AAG GAA GTT CCT CCA AGG ATA AAA TCA AGA ACG AAC CAA    650

 I   K   I   T   F   K   S   D   D   Y   F   V   A   K   P   G   F   K   I   Y    166
ATT AAA ATC ACA TTC AAG TCC GAT GAC TAC TTT GTG GCT AAA CCT GGA TTC AAG ATT TAT    710

 Y   S   L   L   E   D   F   Q   P   A   A   A   S   E   T   N   W   E   S   V    186
TAT TCT TTG CTG GAA GAT TTC CAA CCC GCA GCA GCT TCA GAG ACC AAC TGG GAA TCT GTC    770

 T   S   S   I   S   G   V   S   Y   N   S   P   S   V   T   D   P   T   L   I    206
ACA AGC TCT ATT TCA GGG GTA TCC TAT AAC TCT CCA TCA GTA ACG GAT CCC ACT CTG ATT    830

 A   D   A   L   D   K   K   I   A   E   F   D   T   V   E   D   L   L   K   Y    226
GCG GAT GCT CTG GAC AAA AAA ATT GCA GAA TTT GAT ACA GTG GAA GAT CTG CTC AAG TAC    890

 F   N   P   E   S   W   Q   E   D   L   E   N   M   Y   L   D   T   P   R   Y    246
TTC AAT CCA GAG TCA TGG CAA GAA GAT CTT GAG AAT ATG TAT CTG GAC ACC CCT CGG TAT    950

 R   G   R   S   Y   H   D   R   K   S   K   V   D   L   D   R   L   N   D   D    266
CGA GGC AGG TCA TAC CAT GAC CGG AAG TCA AAA GTT GAC CTG GAT AGG CTC AAT GAT GAT   1010

 A   K   R   Y   S   C   T   P   R   N   Y   S   V   N   I   R   E   E   L   K    286
GCC AAG CGT TAC AGT TGC ACT CCC AGG AAT TAC TCG GTC AAT ATA AGA GAA GAG CTG AAG   1070

 L   A   N   V   V   F   F   P   R   C   L   L   V   Q   R   C   G   G   N   C    306
TTG GCC AAT GTG GTC TTC TTT CCA CGT TGC CTC CTC GTG CAG CGC TGT GGA GGA AAT TGT   1130

 G   C   G   T   V   N   W   R   S   C   T   C   N   S   G   K   T   V   K   K    326
```

FIGURE 1

```
GGC TGT GGA ACT GTC AAC TGG AGG TCC TGC ACA TGC AAT TCA GGG AAA ACC GTG AAA AAG  1190

TAT CAT GAG GTA TTA CAG TTT GAG CCT GGC CAC ATC AAG AGG AGG GGT AGA GCT AAG ACC  1250

 M   A   L   V   D   I   Q   L   D   H   H   E   R   C   D   C   I   C   S   S   366
ATG GCT CTA GTT GAC ATC CAG TTG GAT CAC CAT GAA CGA TGC GAT TGT ATC TGC AGC TCA  1310

 R   P   P   R   *                                                               371
AGA CCA CCT CGA TAA                                                             1325

GAGAATGTGCACATCCTTACATTAAGCCTGAAAGAACCTTTAGTTTAAGGAGGGTGAGATAAGAGACCCTTTTCCTACC  1404

AGCAACCAAACTTACTACTAGCCTGCAATGCAATGAACACAAGTGGTTGCTGAGTCTCAGCCTTGCTTTGTTAATGCCA  1483

TGGCAAGTAGAAAGGTATATCATCAACTTCTATACCTAAGAATATAGGATTGCATTTAATAATAGTGTTTGAGGTTATA  1562

TATGCACAAACACACACAGAAATATATTCATGTCTATGTGTATATAGATCAAATGTTTTTTTTTTGGTATATATAACCA  1641

GGTACACCAGAGGTTACATATGTTTGAGTTAGACTCTTAAAATCCTTTGCCAAAATAAGGGATGGTCAAATATATGAAA  1720

CATGTCTTTAGAAAATTTAGGAGATAAATTTATTTTTAAATTTTGAAACACGAAACAATTTTGAATCTTGCTCTCTTAA  1799

AGAAAGCATCTTGTATATTAAAAATCAAAAGATGAGGCTTTCTTACATATACATCTTAGTTGATTATTAAAAAAGGAAA  1878

AATATGGTTTCCAGAGAAAAGGCCAATACCTAAGCATTTTTTCCATGAGAAGCACTGCATACTTACCTATGTGGACTGT  1957

AATAACCTGTCTCCAAAACCATGCCATAATAATATAAGTGCTTTAGAAATTAAATCATTGTGTTTTTTATGCATTTTGC  2036

TGAGGCATCCTTATTCATTTAACACCTATCTCAAAAACTTACTTAGAAGGTTTTTTATTATAGTCCTACAAAAGACAAT  2115

GTATAAGCTGTAACAGAATTTTGAATTGTTTTTCTTTGCAAAACCCCTCCACAAAAGCAAATCCTTTCAAGAATGGCAT  2194

GGGCATTCTGTATGAACCTTTCCAGATGGTGTTCAGTGAAAGATGTGGGTAGTTGAGAACTTAAAAAGTGAACATTGAA  2273

ACATCGACGTAACTGGAAATTAGGTGGGATATTTGATAGGATCCATATCTAATAATGGATTCGAACTCTCCAAACTACA  2352

CCAATTAATTTAATGTATCTTGCTTTTGTGTTCCCGTCTTTTTGAAATATAGACATGGATTTATAATGGCATTTTATAT  2431

TTGGCAGGCCATCATAGATTATTTACAACCTAAAAGCTTTTGTGTATCAAAAAAATCACATTTTATTAATGTAAATTTC  2510

TAATCGTATACTTGCTCACTGTTCTGATTTCCTGTTTCTGAACCAAGTAAAATCAGTCCTAGAGGCTATGGTTCTTAAT  2589

CTATGGAGCTTGCTTTAAGAAGCCAGTTGTCAATTGTGGTAACACAAGTTTGGCCCTGCTGTCCTACTGTTTAATAGAA  2668

AACTGTTTTACATTGGTTAATGGTATTTAGAGTAATTTTTTCTCTCTGCCTCCTTCGTGTCTGTTTTAAAGGAGACTAA  2747

CTCCAGGAGTAGGAAATGATTCATCATCCTCCAAAGCAAGAGGCTTAAGAGAGAAACACCGAAATTCAGATAGCTCAGG  2826

GACTGCTAACAGAGAACTACATTTTTCTTATTGCCTTGAAAGTTAAAAGGAAAGCAGATTTCTTCAGTGACTTTGTGGT  2905

CCTACTAACTACAACCAGTTTGGGTGACAGGGCTGGTAAAGTCCCAGTGTTAGATGAGTGACCTAAATATACTTAGATT  2984

TCTAAGTATGGTGCTCTCAGGTCCAAGTTCAACTATTCTTAAGCAGTGCAATTCTTCCCAGTTATTTGAGATGAAAGAT  3063
```

FIGURE 1 CONT.

CTCTGCTTATTGAAGATGTACCTTCTAAAACTTTCCTAAAAGTGTCTGATGTTTTTACTCAAGAGGGGAGTGGTAAAAT 3142

TAAATACTCTATTGTTCAATTCTCTAAAATCCCAGAACACAATCAGAAATAGCTCAGGCAGACACTAATAATTAAGAAC 3221

GCTCTTCCTCTTCATAACTGCTTTGCAAGTTTCCTGTGAAAACATCAGTTTCCTGTACCAAAGTCAAAATGAACGTTAC 3300

ATCACTCTAACCTGAACAGCTCACAATGTAGCTGTAAATATAAAAAAATGAGAGTGTTCTACCCAGTTTTCAATAAACCT 3379

TCCAGGCTGCAATAACCAGCAAGGTTTTCAGTTAAAGCCCTATCTGCACTTTTTATTTATTAGCTGAAATGTAAGCAGG 3458

CATATTCACTCACTTTTCTTTGCCTTTCCTGAGAGTTTTATTAAAACTTCTCCCTTGGTTACCTGTTATCTTTTGCACT 3537

TCTAACATGTAGCCAATAAATCTATTTGATAGCCATCAAAGGAATAAAAAGCTGGCCATACAAATTACATTTCAAAACA 3616

AACCCTAATAAATCCACATTTCCGCATGGCTCATTCACCTGGAATAATGCCTTTTATTGAATATGTTCTTATAGGGCAA 3695

AACACTTTCATAAGTAGAGTTTTTTATGTTTTTTGTCATATCGGTAACATGCAGCTTTTTCCTCTCATAGCATTTTCTA 3774

TAGCGAATGTAATATGCCTCTTATCTTCATGAAAAATAAATATTGCTTTTGAACAAAAAAAAAAAAAAAAGGGCGGCCGC 3853

TAGCGAATGTAATATGCCTCTTATCTTCATGAAAAATAAATATTGCTTTTGAACAAAAAAAAAAAAAAAGGGCGGCCGC 3853

FIGURE 1 CONT.

Protein Family / Domain Matches, HMMer version 2

```
Searching for complete domains
hmmpfam - search a single seq against HMM database
HMMER 2.1.1 (Dec 1998)
Copyright (C) 1992-1998 Washington University School of Medicine
HMMER is freely distributed under the GNU General Public License (GPL)
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
HMM file:                 /prod/ddm/seqanal/PFAM/pfam3.4/Pfam
Sequence file:            /tmp/orfanal.15446.aa
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
 Query:  sequence15443

Scores for sequence family classification (score includes all domains):
Model    Description                                        Score    E-value  N
-------- -----------                                        -----    ------- ---
CUB      PF00431 CUB domain                                  81.7    1.5e-20   1


Parsed for domains:
Model    Domain  seq-f seq-t    hmm-f hmm-t       score  E-value
-------- ------- ----- -----    ----- -----       -----  -------
CUB        1/1      53   167 ..     1   116 []    81.7  1.5e-20


Alignments of top-scoring domains:
CUB: domain 1 of 1, from 53 to 167: score 81.7, E = 1.5e-20
                     *->CGgtldltessGsisSPnYPnrsdYppnkeCvWrIrappgyrvVeLt
                          +t+    + +G+ +SP +Pn  +Yp+n+  +Wr+  +++ r ++L
  sequence15     53     -DETI-QVKGNGYVQSPRFPN--SYPRNLLLTWRLHSQENTR-IQLV 94

                     Fqd.FdlEdhdgapCryDyvEirDGdpss.pllGrfCG.sgkPedirSts
                     F+ +F lE+ ++  CryD+vE+ D  + s+ ++Gr CG+++ P+ i+S +
  sequence15     95 FDNqFGLEEAENDICRYDFVEVEDISETStIIRGRWCGhKEVPPRIKSRT 144

                     nrmlikFvs.DasvskrGFkAty<-*
                     n+  i+F+s+D  v+k GFk++y
  sequence15    145 NQIKITFKSdDYFVAKPGFKIYY    167

//
```

FIGURE 2

```
BESTFIT of: vegf-g.pep  check: 429  from: 1  to: 370

EGF-G
                VEGF-C
to:             human.pep  check: 6807  from: 1  to: 419

    _human VEGF-C

Symbol comparison table: /usr/local/gcg_9.1/gcgcore/data/rundata/blosum62.cmp
CompCheck: 6430

        Gap Weight:       5      Average Match:  2.912
     Length Weight:       2   Average Mismatch: -2.003

           Quality:     114             Length:     208
             Ratio:   0.610               Gaps:      15
Percent Similarity:  40.476   Percent Identity:  30.357

       Match display thresholds for the alignment(s):
                    | = IDENTITY
                    : =    2
                    . =    1

vegf-g.pep x vegc_human.pep

    175 PAAASETNWESVTSSISGVSYNSPSVTDPTLIADA.LDKKIAEFDTVEDL 223   VEGF-G
        ||||.   .|   | :  .|   |   : |  |   |:..:    .|::|
     26 PAAAAA..FE...SGLD.LSDAEPDAGEATAYASKDLEEQLRSVSSVDEL 69    VEGF-C

    224 LKYFNPE...........SWQEDLENMYL.....DTPRYRGRSYHDRKS 256
        :   ||             || . |   |     :| ::    |.   .
     70 MTVLYPEYWKMYKCQLRKGGWQHNREQANLNSRTEETIKFAAAHYN...T 116

    257 KVDLDRLNDDAKRYSCTPRNYSVNIREELKLANVVFF.PRCLLVQRCGGN 305
        .: |  :..: ::  | ||   :.: .|  .|   || | |. | ||||
    117 EI.LKSIDNEWRKTQCMPREVCIDVGKEFGVATNTFFKPPCVSVYRCGG. 164

    306 CGCGTVNWRSCTCNSGKTVKKYHEVLQFEPGHIKRRGRAKTMALVDIQLD 355
        | |   |    | .  |   |     ||      .|  |    | |
    165 C.C...NSEGLQCMNTST..SYLSKTLFEITVPLSQG.PKP...VTISFA 204

    356 HHERCDCI 363
        .|  | |.
    205 NHTSCRCM 212
```

FIGURE 3

```
 BESTFIT of: vegf-g.pep  check: 429  from: 1  to: 370

VEGF-G
          VEGF-A
 to:      human.pep  check: 4840  from: 1  to: 215

    human VEGF-A

 Symbol comparison table: /usr/local/gcg_9.1/gcgcore/data/rundata/blosum62.cmp
 CompCheck: 6430

        Gap Weight:       5      Average Match:  2.912
     Length Weight:       2   Average Mismatch: -2.003

           Quality:      71             Length:    116
             Ratio:   0.732               Gaps:      9
 Percent Similarity: 39.583   Percent Identity: 26.042

       Match display thresholds for the alignment(s):
                   | = IDENTITY
                   : =    2
                   . =    1

 vegf-g.pep x vegf_human.pep

     248 GRSYHDRKSKVDLDRLNDDAKRYSCTPRNYSVNIREEL.KLANVVFPPRC 296  VEGF-G
         |..:|      :. :  |  .|  | |     |.| :|       :| | |
      34 GQNHH......EVVKFMDVYQRSYCHPIETLVDIFQEYPDEIEYIFKPSC 77   VEGF-A

     297 LLVQRCGGNCGCGTVNWRSCTCNSGKTVKKYHEVLQFEPGHIKRRGRAKT 346
         . . |||| | |       |        :    ::::. .| |   .|.
      78 VPLMRCGG.C.CND.EGLECVPTEESNITM..QIMRIKP.H...QGQ... 115

     347 MALVDIQLDHHERCDC 362
           : :.    | :|:|
     116 .HIGEMSFLQHNKCEC 130
```

FIGURE 4

```
 BESTFIT of: vegf-g.pep  check: 429  from: 1  to: 370

VEGF-G

 to: t128.pep  check: 7365  from: 1  to: 346

T128

 Symbol comparison table: /usr/local/gcg_9.1/gcgcore/data/rundata/blosum62.cmp
 CompCheck: 6430

        Gap Weight:      12      Average Match:  2.912
     Length Weight:       4   Average Mismatch: -2.003

           Quality:     628             Length:     318
             Ratio:   2.129              Gaps:       8
 Percent Similarity: 60.825   Percent Identity: 51.203

       Match display thresholds for the alignment(s):
                   | = IDENTITY
                   : =    2
                   . =    1

 vegf-g.pep x t128.pep

      51 RRDETIQVKGNGYVQSPRFPNSYPRNLLLTWRLHS.QENTRIQLVFDNQF 99  VEGF-G
         . :  | |  || : |||||..|||| .| ||| . :||  ||| || .|
      45 QHERIITVSTNGSIHSPRFPHTYPRNTVLVWRLVAVEENVWIQLTFDERF 94  T128

     100 GLEEAENDICRYDFVEVEDISETSTIIRGRWCGHKEVPPRIKSRTNQIKI 149
         |||: |.|||:|||||||: |: .  | |||||   || :  |: |||:|
      95 GLEDPEDDICKYDFVEVEEPSDGT..ILGRWCGSGTVPGKQISKGNQIRI 142

     150 TFKSDDYFVAKPGFKIYYSLLEDFQPAAASETNWESVTSSISGVSYNSPS 199
          | ||:|| ..||| |:|.:.                | .:      |||
     143 RFVSDEYFPSEPGFCIHYNIV...........MPQFTEAV.....SPS 174

     200 VTDPT.LIADALDKKIAEFDTVEDLLKYFNPESWQEDLENMYLDTPRYRG 248
         |  |. |  | |.  |  | |.|||::|  || || |||.:|  | .  |
     175 VLPPSALPLDLLNNAITAFSTLEDLIRYLEPERWQLDLEDLYRPTWQLLG 224

     249 RSY.HDRKSK.VDLDRLNDDAKRYSCTPRNYSVNIREELKLANVVFFPRC 296
         :.:   |||: |||. | :: : |||||||:||.||||||  . :|.| |
     225 KAFVFGRKSRVVDLNLLTEEVRLYSCTPRNFSVSIREELKRTDTIFWPGC 274

     297 LLVQRCGGNCGCGTVNWRSCTCNSGKTVKKYHEVLQFEPGHIKRRGRAKT 346
         |||.||||||  |   |   | |   |  ||||||||  |   |   |
     275 LLVKRCGGNCACCLHNCNECQCVPSKVTKKYHEVLQLRP...KTGVRGLH 321

     347 MALVDIQLDHHERCDCIC 364
          .| |: |:||| |||:|
     322 KSLTDVALEHHEECDCVC 339
```

FIGURE 5

```
ALIGN calculates a global alignment of two sequences
 version 2.0uPlease cite: Myers and Miller, CABIOS (1989)
, VEGF-G                                          370 aa vs.
, T128 a.a                                        345 aa
scoring matrix: pam120.mat, gap penalties: -12/-4
42.4% identity;          Global alignment score: 611

               10        20        30        40        50        60        70
inputs MHRLIFVYTLICANFCSCRDTSATPQSASIKALRNANLRRDESNHLTDLYRRDETIQVKGNGYVQSPRFP  VEGF-G
       :  .:   :. ... . :. . .  . : :   ..:    : : . :  ... .: :  :: .:::::::
       MS--LFGLLLLTSALAGQRQGTQAESNLSSKFQFSSN---KEQNGVQDP-QHERIITVSTNGSIHSPRFP  T128
                 10        20        30            40          50        60

               80        90       100       110       120       130
inputs NSYPRNLLLTWRLHS-QENTRIQLVFDNQFGLEEAENDICRYDFVEVEDISETSTIIRGRWCGHKEVPPR
       ..:::: .:.::: . .:::.:::::.::..:::::.:::::::::::::::. :. .:::. :::::   :: .
       HTYPRNTVLVWRLVAVEENVWIQLTFDERFGLEDPEDDICKYDFVEVEEPSD-GTIL-GRWCGSGTVPGK
           70        80        90       100       110       120       130

     140       150       160       170       180       190       200
inputs IKSRTNQIKITFKSDDYFVAKPGFKIYYSLLEDFQPAAASETNWESVTSSISGVSYNSPSVTDPTLIADA
       :. :::.: : :::.:: . ::: : :...    :  .     :.:...:.          .  ..:  :
       QISKGNQIRIRFVSDEYFPSEPGFCIHYNIV---MPQFT-----EAVSPSV---------LPPSALPLDL
            140       150       160           170                 180

     210       220       230       240       250       260       270
inputs LDKKIAEFDTVEDLLKYFNPESWQEDLENMYLDTPRYRGRSY-HDRKSKV-DLDRLNDDAKRYSCTPRNY
       :.. :...:.:::::.:::: :: :::::.:  : .  :...  .::::.: ::. :..... ::::::::.
       LNNAITAFSTLEDLIRYLEPERWQLDLEDLYRPTWQLLGKAFVFGRKSRVVDLNLLTEEVRLYSCTPRNF
         190       200       210       220       230       240       250

       280       290       300       310       320       330       340
inputs SVNIREELKLANVVFFPRCLLVQRCGGNCGCGTVNWRSCTCNSGKTVKKYHEVLQFEPGHIKRRGRAKTM
       ::.::::::: ....: : ::::.:::::::.:    :   : : ..:..:::::::::. :   :   :.  .
       SVSIREELKRTDTIFWPGCLLVKRCGGNCACCLHNCNECQCVPSKVTKKYHEVLQLRP---KTGVRGLHK
         260       270       280       290       300       310       320

       350       360       370
inputs ALVDIQLDHHERCDCICSSRPPR
       .:.:. :.::: :::.: .
       SLTDVALEHHEECDCVCRGSTGG
            330       340
```

FIGURE 6

```
ALIGN calculates a global alignment of two sequences
 version 2.0uPlease cite: Myers and Miller, CABIOS (1989)
                VEGF-G                                  1110 aa vs
                                                        1035 aa
scoring matrix: pam120.mat, gap penalties: 12/-4
51.9% identity;          Global alignment score: 1798

                10         20         30         40         50         60
inputs ATGCACCGGCTCAT-CTTTGTCTACACTCTAATCTGCGCAAACTTTTGCAGCTGTC--GGGACACTT-CT    VEGF-G
       ::: .::   ::.  :::   :::::. ::  :::::: :....    :: . .:.:  ::::::::  :
       ATGAGCCTCTTCGGGCTTCTCCTGCTGAC--ATCTGCCCTGGCCG--GCCAGAGACAGGGGACTCAGGCG    7128
                10         20         30         40         50         60

          70         80         90        100        110        120        130
inputs GCAACCCCGCAGAGCGCATCCATCAAAGCTTTGCGCAACGCCAACCTCAGGCGAGA-TGAGAGCAATCAC
       : :.::   :.::: . ..  .:: :.  ::: : :::.. :::    ...:...:  :::.:::  .:
       GAATCCAACCTGAGTAGTAAATTCCAG--TTTTC-CAGCAACAAG---GAACAGAACGGAGTACAAGATC
          70         80         90        100        110        120        130

          140        150        160        170        180        190        200
inputs CTCA-CAGACTTGTACCGAAGAGATGAGACCATCCAGGTGAAAGGAAACGGCTACGTGCAGAGTCCTAGA
       :::: :: .  .:.:   .: .. ::.:.:  : :.....:::: :.... :.::          :::.:.
       CTCAGCATGAGAGAA--TTATTACTGTGTC--TACTAATGGAAGTATTCA-CAGC---------CCAAGG
              140        150        160        170        180

          210        220        230        240        250        260        270
inputs TTCCCGAACAGCTACCCCAGGAACCTGCTCCTGACATGGCGGCT--TCACTCTC-AGGAGAATACACGGA
       :: ::  : :  :: :: ::.::   : :: ::. :::: :. :  : .:: :  :::::::  : :::
       TTTCCTCATACTTATCCAAGAAATACGGTCTTGGTATGGAGATTAGTAGCAGTAGAGGAAAATGTATGGA
        190        200        210        220        230        240        250

                280        290        300        310        320        330        340
inputs TACAGCTAGTGTTTGACAATCAGTTTGGATTAGAGGAAGCAGAAAATGATATCTGTAGGTATGATTTTGT
       :::.:::.. :::::: .:. .:.::::: :.:::.::  ::::::.:::: :: :: :.:::::::::::::
       TACAACTTACGTTTGATGAAAGATTTGGGCTTGAAGACCCAGAAGATGACATATGCAAGTATGATTTTGT
        260        270        280        290        300        310        320

                350        360        370        380        390        400        410
inputs GGAAGTTGAAGATATATCCGAAACCAGTACCATTATTAGAGGACGATGGTGTGGACACAAGGA-AGTTCC
       .::::::::.::    : ::.... .:.:::   ::::. :::.:: ::::::::. .:.:: : .:::.::
       AGAAGTTGAGGA---ACCCAGTGATGGAAC---TATATTAGGGCGCTGGTGTGGT-TCTGGTACTGTACC
        330            340        350            360        370        380

                420        430        440        450        460        470        480
inputs TCCAAGGATAAAATCAAGAACGAACCAAATTAAAATCACATTCAAGTCCGATGACTACTTTGTGGCTAAA
       .  ::.. ..:.::.:.:. .:: :::::::::.:: : ::: ...:: ::::: :: :::    ::.::
       AGGAAAACAGATTTCTAAAGGAAATCAAATTAGGATAAGATTTGTATCTGATGAATATTTTCCTTCTGAA
     390        400        410        420        430        440        450

                490        500        510        520        530        540        550
inputs CCTGGATTCAAGATTTATTATTCTTTGCTGGAAGATTTCCAACCCGCAGCAGCTTCAGAGACCAACTGGG
       ::.::.::::..          :: :. :.:   :: :::  :::. ::.::.    :::: :::  :.::: :
       CCAGGGTTCTG-------CATCCACTAC---AACATTGTCATGCCACAA----TTCACAGA--AGCTGTG
     460        470                480        490            500        510

                560        570        580        590        600        610        620
inputs AATCTGTCACAAGCTCTATTTCAGGGGTATCCTATAACTCTCCATCAGTAACGGATCCCACTCTGATTGC
       :.::          ::     :::: : :: ::          ::.:::::   :  . ::   .::::.. :
       AGTC---------CT-----TCAGTGCTACCC---------CCTTCAG---CTTTGCC---ACTGGAC-C
```

FIGURE 7

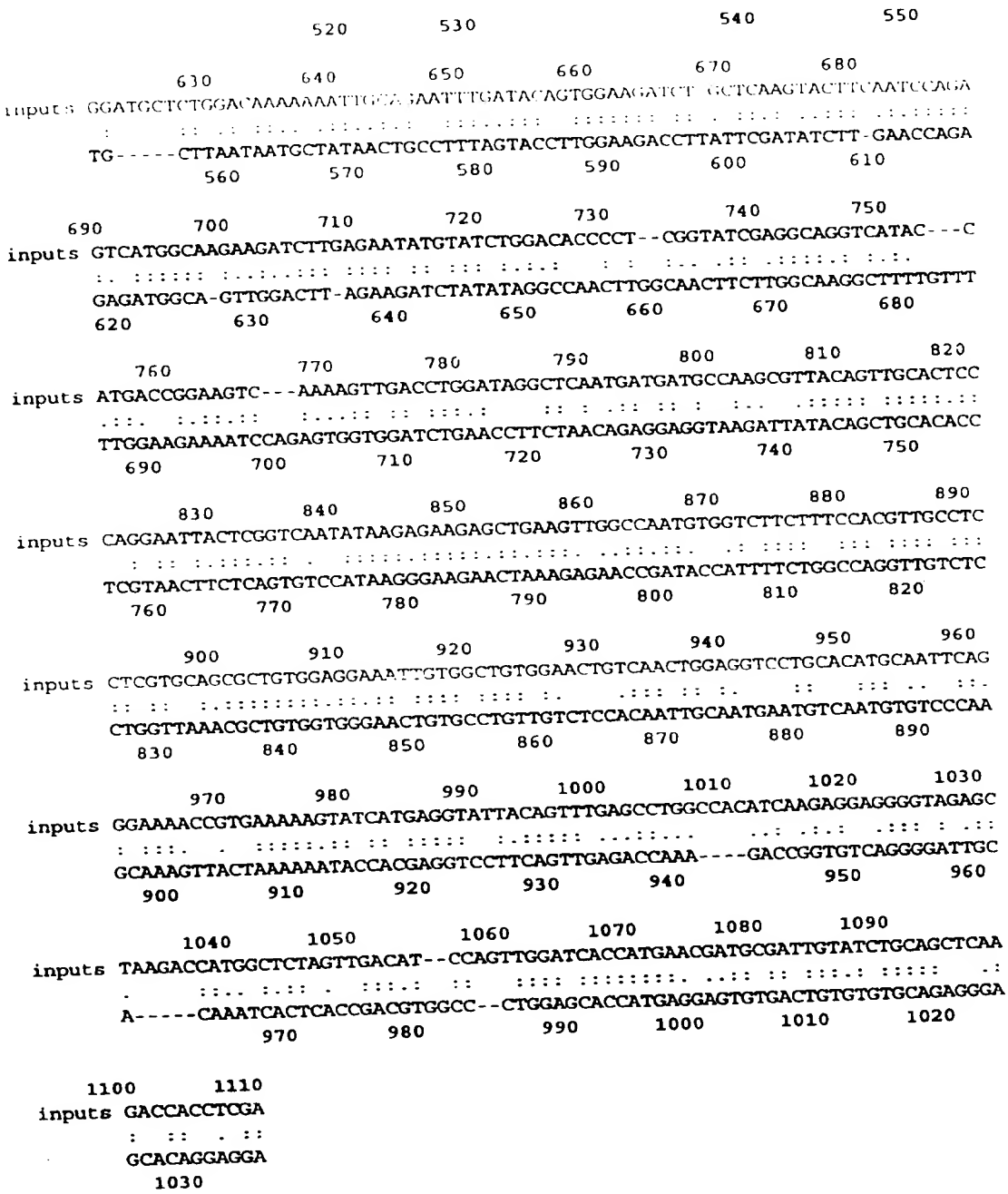

FIGURE 7 CONT.

```
gaattcccgg gtcgacccac gcgtccgagc gccagtaact gggcggccgg gacaacacag      60
gcggtgaagg cgagggactg tgcagtagaa atccgccgac tcaacccttt gggctttatt     120
tatttacttt tggagcaacg cgatccctag gtcgctgagc ccaa atg caa cgg ctc      176
                                                 Met Gln Arg Leu
                                                   1

gtt tta gtc tcc att ctc ctg tgc gcg aac ttt agc tgc tat ccg gac       224
Val Leu Val Ser Ile Leu Leu Cys Ala Asn Phe Ser Cys Tyr Pro Asp
 5                  10                  15                  20

act ttt gcg act ccg cag aga gca tcc atc aaa gct ttg cgc aat gcc       272
Thr Phe Ala Thr Pro Gln Arg Ala Ser Ile Lys Ala Leu Arg Asn Ala
                25                  30                  35

aac ctc agg aga gat gag agc aat cac ctc aca gac ttg tac cag aga       320
Asn Leu Arg Arg Asp Glu Ser Asn His Leu Thr Asp Leu Tyr Gln Arg
            40                  45                  50

gag gag aac att cag gtg aca agc aat ggc cat gtg cag agt cct cgc       368
Glu Glu Asn Ile Gln Val Thr Ser Asn Gly His Val Gln Ser Pro Arg
        55                  60                  65

ttc ccg aac agc tac cca agg aac ctg ctt ctg aca tgg tgg ctc cgt       416
Phe Pro Asn Ser Tyr Pro Arg Asn Leu Leu Leu Thr Trp Trp Leu Arg
    70                  75                  80

tcc cag gag aaa aca cgg ata caa ctg tcc ttt gac cat caa ttc gga       464
Ser Gln Glu Lys Thr Arg Ile Gln Leu Ser Phe Asp His Gln Phe Gly
 85                  90                  95                 100

cta gag gaa gca gaa aat gac att tgt agg tat gac ttt gtg gaa gtt       512
Leu Glu Glu Ala Glu Asn Asp Ile Cys Arg Tyr Asp Phe Val Glu Val
                105                 110                 115

gaa gaa gtc tca gag agc agc act gtt gtc aga gga aga tgg tgt ggc       560
Glu Glu Val Ser Glu Ser Ser Thr Val Val Arg Gly Arg Trp Cys Gly
            120                 125                 130

cac aag gag atc cct cca agg ata acg tca aga aca aac cag att aaa       608
His Lys Glu Ile Pro Pro Arg Ile Thr Ser Arg Thr Asn Gln Ile Lys
        135                 140                 145

atc aca ttt aag tct gat gac tac ttt gtg gca aaa cct gga ttc aag       656
Ile Thr Phe Lys Ser Asp Asp Tyr Phe Val Ala Lys Pro Gly Phe Lys
    150                 155                 160

att tat tat tca ttt gtg gaa gat ttc caa ccg gaa gca gcc tca gag       704
Ile Tyr Tyr Ser Phe Val Glu Asp Phe Gln Pro Glu Ala Ala Ser Glu
165                 170                 175                 180

acc aac tgg gaa tca gtc aca agc tct ttc tct ggg gtg tcc tat cac       752
Thr Asn Trp Glu Ser Val Thr Ser Ser Phe Ser Gly Val Ser Tyr His
                185                 190                 195
```

FIGURE 8

```
tct cca tca ata acg gac ccc act ctc act gct gat gcc ctg gac aaa      800
Ser Pro Ser Ile Thr Asp Pro Thr Leu Thr Ala Asp Ala Leu Asp Lys
            200                 205                 210

act gtc gca gaa ttc gat acc gtg gaa gat cta ctt aag cac ttc aat      848
Thr Val Ala Glu Phe Asp Thr Val Glu Asp Leu Leu Lys His Phe Asn
        215                 220                 225

cca gtg tct tgg caa gat gat ctg gag aat ttg tat ctg gac acc cct      896
Pro Val Ser Trp Gln Asp Asp Leu Glu Asn Leu Tyr Leu Asp Thr Pro
    230                 235                 240

cat tat aga ggc agg tca tac cat gat cgg aag tcc aaa gtg gac ctg      944
His Tyr Arg Gly Arg Ser Tyr His Asp Arg Lys Ser Lys Val Asp Leu
245                 250                 255                 260

gac agg ctc aat gat gat gtc aag cgt tac agt tgc act ccc agg aat      992
Asp Arg Leu Asn Asp Asp Val Lys Arg Tyr Ser Cys Thr Pro Arg Asn
                265                 270                 275

cac tct gtg aac ctc agg gag gag ctg aag ctg acc aat gca gtc ttc     1040
His Ser Val Asn Leu Arg Glu Glu Leu Lys Leu Thr Asn Ala Val Phe
            280                 285                 290

ttc cca cga tgc ctc ctc gtg cag cgc tgt ggt ggc aac tgt ggt tgc     1088
Phe Pro Arg Cys Leu Leu Val Gln Arg Cys Gly Gly Asn Cys Gly Cys
        295                 300                 305

gga act gtc aac tgg aag tcc tgc aca tgc agc tca ggg aag aca gtg     1136
Gly Thr Val Asn Trp Lys Ser Cys Thr Cys Ser Ser Gly Lys Thr Val
    310                 315                 320

aag aag tat cac gag gta ttg aag ttt gag cct gga cat ttc aag aga     1184
Lys Lys Tyr His Glu Val Leu Lys Phe Glu Pro Gly His Phe Lys Arg
325                 330                 335                 340

agg ggc aaa gct aag aat atg cct ctt gtt gat atc cag ctg gat cat     1232
Arg Gly Lys Ala Lys Asn Met Pro Leu Val Asp Ile Gln Leu Asp His
                345                 350                 355

cat gag cga tgt gac tgt atc tgc agc tca aga cca cct cga             1274
His Glu Arg Cys Asp Cys Ile Cys Ser Ser Arg Pro Pro Arg
            360                 365                 370

taaaacacta tgcacatctg tactttgatt atgaaaggac ctttaggtta caaaaaccct   1334
aagaagcttc taatctcagt gcaatgaatg catatggaaa tgttgctttg ttagtgccat   1394
ggcaagaaga agcaaatatc attaatttct atatacataa acataggaat tcacttatca   1454
atagtatgtg aagatatgta tatatactta tatacatgac tagctctatg tatgtaaata   1514
gattaaatac tttattcagt atatttactg aagtccccta gaggcaaatg atattacttg   1574
atattttgct aaaataaggg aatgtgtcaa gcatataaaa tatctttcaa aaattcaaaa   1634
ggtaagttta ttctaaagct ttgtatgaca aaatatatca gattttgctc acttaaagaa   1694
ggcacttgtc cattaaatga aagatgagac tttcttctgt gtataaatct aattatgtga   1754
agacaaatac atagtttctg gcaataaaaa caagttttga aatattttct tcatgagatg   1814
tactactcac ctaccaatgt ggacagttat tatctgtctc cacaactatg ccaaagtaat   1874
acaggatatt taacaatcaa gtcaatcaag tcattaagtc ctcttccatg tattttctca   1934
atgcaagcta agtaactgag cacccttctc aatgaattgc tcagggactt gtggtgataa   1994
ttggcaagag atattcaata ggcaagcaat aggttgctgt agaatttttg aatttttttt   2054
```

FIGURE 8, CONT.

```
tattttaaag tctgcagaaa agcaagtgtt ttcaggaaga acatgatatt tattacacat  2114
gagcctttaa atatggtagt catggcaaga cttggggtgat agaaagtcag aaataagcaa  2174
atgttgacat aactggaaat caaatgggct gtttgataag gtcttacatt aaaagtgggt  2234
caaaaatctt ctaaactgtg ccaataattt ttctgtgttt tgcttcttgca ttcctatctt  2294
ttcaatatat acacatgaac ttatattgtg ctatatgaga caagctatca tggaatattt  2354
atagcctaaa atatttatgt attaacaaaa caactagagc cagctaagtt gtggaggcac  2414
agatttggtc ctgctgttcc actggaaagc ttttgttcct gatttttagg agtgagttaa  2474
gtttccccag tggtgcctct atatctttcc taaagagatt gagtccagaa acaggaagtt  2534
gctcattgtt ctccaaacaa aatgtttagg agggaatcct gaagtcaagt taacttggtt  2594
ctgctactag gaaactatat ttttcttcct gccttttagag tagaaaagaa aactgattta  2654
ttcggtgagt ccaacactac cctgagcaca aagtgtttg tttctgaatc tgtgcatagg  2714
ccctttttgt ctgatgtcag atgaatagct tatacacact tagattgcta aggatgatgc  2774
tgtctaaacc aagttatcag ctgtgtgctc aacaaatctg tgcttactga cacatatatt  2834
ctaaaggttc tatgcagtca attcctactt aaatgattat taaaattaaa  2894
tagtatattg ttgtcaagca tgccacagaa caatgagaag taagccaggc ctaatttgca  2954
agcttcctgt gaaacaatgg ctcctataaa aaaatcaaaa tgtacattac cttactcact  3014
gtaaaggaca cagatcaatc ataactgtaa aaatgaaatg ttaagactgt tttttcaata  3074
aaccttctag gtggtgataa ccaaaaaaaa aaaaaaaggg cggccgc  3121
```

FIGURE 8, CONT.

```
hmmpfam - search a single seq against HMM database
HMMER 2.1.1 (Dec 1998)
Copyright (C) 1992-1998 Washington University School of Medicine
HMMER is freely distributed under the GNU General Public License (GPL).
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
HMM file:                 /prod/ddm/seqanal/PFAM/pfam5.2/Pfam
Sequence file:            /prod/ddm/wspace/orfanal/oa-script.3303.seq
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
 Query:  mouseT274_VEGF_

Scores for sequence family classification (score includes all domains):
Model         Description                                    Score    E-value  N
--------      -----------                                    -----    ------- ---
CUB            CUB domain                                     90.6    3.2e-23   1


Parsed for domains:
Model         Domain  seq-f seq-t    hmm-f hmm-t       score  E-value
--------      ------- ----- -----    ----- -----       -----  -------
CUB             1/1      53   167 ..     1   116 []    90.6  3.2e-23


Alignments of top-scoring domains:
CUB: domain 1 of 1, from 53 to 167: score 90.6, E = 3.2e-23
                   *->CGgtldltessGsisSPnYPnrsdYppnkeCvWrIrappgyrvVeLt
                        +++   +s+G+ +SP +Pn  +Yp+n+  +W +r ++  r ++L+
  mouseT274_    53    -EENI-QVTSNGHVQSPRFPN--SYPRNLLLTWWLRSQEKTR-IQLS 94

                   Fqd.FdlEdhdgapCryDyvEirDGdpss.pllGrfCG.sgkPedirSts
                   F+++F lE+ ++  CryD+vE+ +  +ss+ ++Gr CG++++P+ i+S +
  mouseT274_    95 FDHqFGLEEAENDICRYDFVEVEEVSESStVVRGRWCGhKEIPPRITSRT 144

                   nrmlikFvs.DasvskrGFkAty<-*
                   n+  i+F+s+D  v+k GFk++y
  mouseT274_   145 NQIKITFKSdDYFVAKPGFKIYY    167
```

FIGURE 9A

```
//
Searching for complete domains in SMART
hmmpfam - search a single seq against HMM database
HMMER 2.1.1 (Dec 1998)
Copyright (C) 1992-1998 Washington University School of Medicine
HMMER is freely distributed under the GNU General Public License (GPL).
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
HMM file:                 /ddm/robison/smart/smart/smart.all.hmms
Sequence file:            /prod/ddm/wspace/orfanal/oa-script.3303.seq
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
 Query:  mouseT274_VEGF_

Scores for sequence family classification (score includes all domains):
Model    Description                                        Score    E-value  N
-------- -----------                                        -----    ------- ---
CUB_2                                                        97.8    2.1e-25   1
PDGF_2                                                        5.2      0.002   1

Parsed for domains:
Model    Domain  seq-f seq-t    hmm-f hmm-t       score  E-value

-------- ------- ----- -----    ----- -----      -----  -------
CUB_2      1/1      54   170 ..     1   144 []    97.8  2.1e-25
PDGF_2     1/1     271   364 ..     1    91 []     5.2    0.002

Alignments of top-scoring domains:
CUB_2: domain 1 of 1, from 54 to 170: score 97.8, E = 2.1e-25
                   *->CGgtltasssDfkesGtitSPnYPnspsgesYpnnleCvWtIsappG
                       +  +t +       G+++SP +Pns     Yp+nl  +W ++  +
  mouseT274_    54     ENIQVTSN-------GHVQSPRFPNS-----YPRNLLLTWWLRSQEK 88

                   yrieLkFtdhdkFdlEssdndgggrfvpeCryDyveiyDGpsktsspllG
                    ri+L+F++   F lE  +nd        CryD+ve+ + ++  ss+++
  mouseT274_    89 TRIQLSFDHQ--FGLEEAENDI-------CRYDFVEVEEVSE--SSTVVR 127

                   ntearfCGse..p.iisSssNsmtvtFvsDssvqgkgktkrGFsarYsav
                       r CG+++ p+ i+S +N++ +tF+sD          k GF+++Ys+v
  mouseT274_   128 G---RWCGHKeiPpRITSRTNQIKITFKSDDYFVA----KPGFKIYYSFV 170

                   <-*

  mouseT274_       -      -

PDGF_2: domain 1 of 1, from 271 to 364: score 5.2, E = 0.002
                   *->seCkPRevvVevseeypdetsAnflfkPpCVtVkRCgGCCNdegLe.
                        C PR   V + ee+ + t  n  f P+C  V+RCgG C+   +
  mouseT274_   271     -SCTPRNHSVNLREEL-KLT--NAVFFPRCLLVQRCGGNCGCGTVNw 313

                   ..CvPtetsnrlVtm..qlleIsr....kkpsssgpelvevsfeeHtkCe
                   ++C+ ++      V++ +++l+  +++ +++++ +++ lv + +  H +C+
  mouseT274_   314 ksCTCSSGKT--VKKyhEVLKFEPghfkRRGKAKNMPLVDIQLDHHERCD 361

                   CrP<-*
                   C+
  mouseT274_   362 CIC     364
```

FIGURE 9B

```
GAP of: FrGcgManager_439_OFD8yopr_  check: 718  from: 1  to: 1110

mVEGFG ORF (analysis only) - Import - complete

to: FrGcgManager_439_PFD0SGVzu  check: 2855  from: 1  to: 1110

h VEGFG ORF (analysis only) - Import - complete

Symbol comparison table: /ddm_local/gcg/gcg_9.1/gcgcore/data/rundata/nwsgapdna.cmp
CompCheck: 8760

        Gap Weight:     12      Average Match: 10.000
     Length Weight:      4   Average Mismatch:  0.000

           Quality:   9500             Length:   1110
             Ratio:  8.559               Gaps:      0
Percent Similarity: 85.586   Percent Identity: 85.586

       Match display thresholds for the alignment(s):
                  | = IDENTITY
                  : =    5
                  . =    1

FrGcgManager_439_OFD8yopr_ x FrGcgManager_439_PFD0SGVzu

      .         .         .         .         .
    1 ATGCAACGGCTCGTTTTAGTCTCCATTCTCCTGTGCGCGAACTTTAGCTG 50    mouse
      ||||| |||||| | || |||| || |||  | ||||| |||||| || |
    1 ATGCACCGGCTCATCTTTGTCTACACTCTAATCTGCGCAAACTTTTGCAG 50    human
      .         .         .         .         .
   51 CTATCCGGACACTTTTGCGACTCCGCAGAGAGCATCCATCAAAGCTTTGC 100
      || || |||||||| ||| || |||||||| ||||||||||||||||||||
   51 CTGTCGGGACACTTCTGCAACCCCGCAGAGCGCATCCATCAAAGCTTTGC 100
      .         .         .         .         .
  101 GCAATGCCAACCTCAGGAGAGATGAGAGCAATCACCTCACAGACTTGTAC 150
      |||| |||||||||||| ||||||||||||||||||||||||||||||||
  101 GCAACGCCAACCTCAGGCGAGATGAGAGCAATCACCTCACAGACTTGTAC 150
      .         .         .         .         .
  151 CAGAGAGAGGAGAACATTCAGGTGACAAGCAATGGCCATGTGCAGAGTCC 200
      |  ||||| |||| ||| ||||||| | | || ||| | ||||||||||
  151 CGAAGAGATGAGACCATCCAGGTGAAAGGAAACGGCTACGTGCAGAGTCC 200
      .         .         .         .         .
  201 TCGCTTCCCGAACAGCTACCCAAGGAACCTGCTTCTGACATGGTGGCTCC 250
      | | ||||||||||||||||| |||||||||| ||||||||| |||| |
  201 TAGATTCCCGAACAGCTACCCCAGGAACCTGCTCCTGACATGGCGGCTTC 250
      .         .         .         .         .
  251 GTTCCCAGGAGAAAACACGGATACAACTGTCCTTTGACCATCAATTCGGA 300
        || |||||||| |||||||||| ||    |||||| |||| || |||
  251 ACTCTCAGGAGAATACACGGATACAGCTAGTGTTTGACAATCAGTTTGGA 300
      .         .         .         .         .
  301 CTAGAGGAAGCAGAAAATGACATTTGTAGGTATGACTTTGTGGAAGTTGA 350
       |||||||||||||||||| || |||||||||| ||||||||||||||
  301 TTAGAGGAAGCAGAAAATGATATCTGTAGGTATGATTTTGTGGAAGTTGA 350
      .         .         .         .         .
  351 AGAAGTCTCAGAGAGCAGCACTGTTGTCAGAGGAAGATGGTGTGGCCACA 400
      |||  | || || | ||| ||  || | |||||| |||||||||| ||||
  351 AGATATATCCGAAACCAGTACCATTATTAGAGGACGATGGTGTGGACACA 400
      .         .         .         .         .
  401 AGGAGATCCCTCCAAGGATAACGTCAAGAACAAACCAGATTAAAATCACA 450
      ||||  | |||||||||||||  |||||||| ||||| ||||||||||||
  401 AGGAAGTTCCTCCAAGGATAAAATCAAGAACGAACCAAATTAAAATCACA 450
```

FIGURE 10

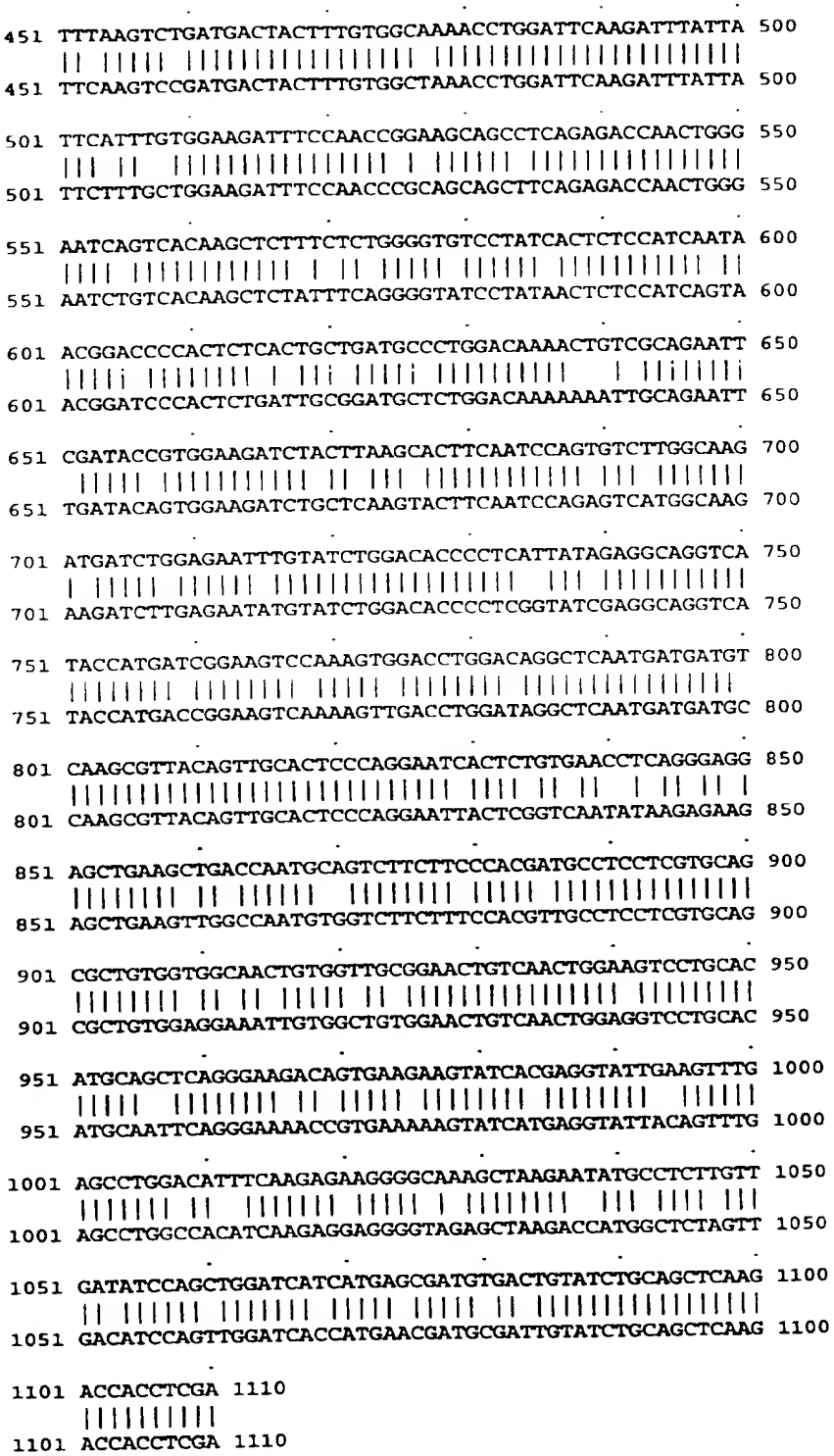

FIGURE 10, CONT.

```
GAP of: FrGcgManager_440_TFD0x.gUZ  check: 9937  from: 1  to: 370

 mVEGFG prot (analysis only) - Import - complete

 to: FrGcgManager_440_UFDxJGDT_  check: 429  from: 1  to: 370

 h VEGFG prot (analysis only) - Import - complete

 Symbol comparison table: /prod/ddm/seqanal/BLAST/matrix/aa/BLOSUM62
 CompCheck: 1102
  Matrix made by matblas from blosum62.iij

        Gap Weight:     12      Average Match:  2.778
     Length Weight:      4   Average Mismatch: -2.248

           Quality:   1732             Length:    370
             Ratio:  4.681               Gaps:      0
 Percent Similarity: 89.730   Percent Identity: 84.865

       Match display thresholds for the alignment(s):
                   | = IDENTITY
                   : =    2
                   . =    1

 FrGcgManager_440_TFD0x.gUZ x FrGcgManager_440_UFDxJGDT_

                .         .         .         .         .
      1 MQRLVLVSILLCANFSCYPDTFATPQRASIKALRNANLRRDESNHLTDLY 50     mouse
        | ||: |  |:||||     || |||| ||||||||||||||||||||||||
      1 MHRLIFVYTLICANFCSCRDTSATPQSASIKALRNANLRRDESNHLTDLY 50     human
                .         .         .         .         .
     51 QREENIQVTSNGHVQSPRFPNSYPRNLLLTWWLRSQEKTRIQLSFDHQFG 100
        .|:| |||  ||:||||||||||||||||| | ||| ||||| ||.|||
     51 RRDETIQVKGNGYVQSPRFPNSYPRNLLLTWRLHSQENTRIQLVFDNQFG 100
                .         .         .         .         .
    101 LEEAENDICRYDFVEVEEVSESSTVVRGRWCGHKEIPPRITSRTNQIKIT 150
        ||||||||||||||||::||.||::|||||||||:|||| |||||||||
    101 LEEAENDICRYDFVEVEDISETSTIIRGRWCGHKEVPPRIKSRTNQIKIT 150
                .         .         .         .         .
    151 FKSDDYFVAKPGFKIYYSFVEDFQPEAASETNWESVTSSPSGVSYHSPSI 200
        |||||||||||||||||| .||||| |||||||||||| |||||.|||:
    151 FKSDDYFVAKPGFKIYYSLLEDFQPAAASETNWESVTSSISGVSYNSPSV 200
                .         .         .         .         .
    201 TDPTLTADALDKTVAEFDTVEDLLKHFNPVSWQDDLENLYLDTPHYRGRS 250
        ||||| |||||| :|||||||||||:||| |||:||||:||||| |||||
    201 TDPTLIADALDKKIAEFDTVEDLLKYFNPESWQEDLENMYLDTPRYRGRS 250
                .         .         .         .         .
    251 YHDRKSKVDLDRLNDDVKRYSCTPRNHSVNLREELKLTNAVFFPRCLLVQ 300
        |||||||||||||||| |||||||||:|||:|||||| | ||||||||||
    251 YHDRKSKVDLDRLNDDAKRYSCTPRNYSVNIREELKLANVVFFPRCLLVQ 300
                .         .         .         .         .
    301 RCGGNCGCGTVNWKSCTCSSGKTVKKYHEVLKFEPGHFKRRGKAKNMPLV 350
        |||||||||||||:||||.|||||||||||||.||||| ||||:|| | ||
    301 RCGGNCGCGTVNWRSCTCNSGKTVKKYHEVLQFEPGHIKRRGRAKTMALV 350
                .         .
    351 DIQLDHHERCDCICSSRPPR 370
        ||||||||||||||||||||
    351 DIQLDHHERCDCICSSRPPR 370
```

FIGURE 11

SEQUENCE LISTING

<110> Millenium Pharmaceuticals, Inc.

<120> A NOVEL VASCULAR ENDOTHELIAL GROWTH FACTOR FAMILY MEMBER AND USES THEREOF

<130> MNI-124CPPC

<140>
<141>

<150> USSN 09/343,671
<151> 1999-06-30

<160> 9

<170> PatentIn Ver. 2.0

<210> 1
<211> 3853
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (213)..(1322)

<400> 1
gtcgacccac gcgtccggga agtgcggcac ccacacgcgc tcggaaagtt cagcatgcag 60

gaagtttggg gagagctcgg cgattagcac agcgacccgg gccagcgcag ggcgagcgca 120

ggcggcgaga gcgcagggcg gcgcggcgtc ggtcccggga gcagaacccg gctttttctt 180

ggagcgacgc tgtctctagt cgctgatccc aa atg cac cgg ctc atc ttt gtc 233
Met His Arg Leu Ile Phe Val
1 5

tac act cta atc tgc gca aac ttt tgc agc tgt cgg gac act tct gca 281
Tyr Thr Leu Ile Cys Ala Asn Phe Cys Ser Cys Arg Asp Thr Ser Ala
10 15 20

acc ccg cag agc gca tcc atc aaa gct ttg cgc aac gcc aac ctc agg 329
Thr Pro Gln Ser Ala Ser Ile Lys Ala Leu Arg Asn Ala Asn Leu Arg
25 30 35

cga gat gag agc aat cac ctc aca gac ttg tac cga aga gat gag acc 377
Arg Asp Glu Ser Asn His Leu Thr Asp Leu Tyr Arg Arg Asp Glu Thr
40 45 50 55

atc cag gtg aaa gga aac ggc tac gtg cag agt cct aga ttc ccg aac 425
Ile Gln Val Lys Gly Asn Gly Tyr Val Gln Ser Pro Arg Phe Pro Asn
60 65 70

agc tac ccc agg aac ctg ctc ctg aca tgg cgg ctt cac tct cag gag 473
Ser Tyr Pro Arg Asn Leu Leu Leu Thr Trp Arg Leu His Ser Gln Glu
75 80 85

```
aat aca cgg ata cag cta gtg ttt gac aat cag ttt gga tta gag gaa   521
Asn Thr Arg Ile Gln Leu Val Phe Asp Asn Gln Phe Gly Leu Glu Glu
         90                  95                  100

gca gaa aat gat atc tgt agg tat gat ttt gtg gaa gtt gaa gat ata   569
Ala Glu Asn Asp Ile Cys Arg Tyr Asp Phe Val Glu Val Glu Asp Ile
    105                 110                 115

tcc gaa acc agt acc att att aga gga cga tgg tgt gga cac aag gaa   617
Ser Glu Thr Ser Thr Ile Ile Arg Gly Arg Trp Cys Gly His Lys Glu
120                 125                 130                 135

gtt cct cca agg ata aaa tca aga acg aac caa att aaa atc aca ttc   665
Val Pro Pro Arg Ile Lys Ser Arg Thr Asn Gln Ile Lys Ile Thr Phe
                140                 145                 150

aag tcc gat gac tac ttt gtg gct aaa cct gga ttc aag att tat tat   713
Lys Ser Asp Asp Tyr Phe Val Ala Lys Pro Gly Phe Lys Ile Tyr Tyr
            155                 160                 165

tct ttg ctg gaa gat ttc caa ccc gca gca gct tca gag acc aac tgg   761
Ser Leu Leu Glu Asp Phe Gln Pro Ala Ala Ala Ser Glu Thr Asn Trp
        170                 175                 180

gaa tct gtc aca agc tct att tca ggg gta tcc tat aac tct cca tca   809
Glu Ser Val Thr Ser Ser Ile Ser Gly Val Ser Tyr Asn Ser Pro Ser
    185                 190                 195

gta acg gat ccc act ctg att gcg gat gct ctg gac aaa aaa att gca   857
Val Thr Asp Pro Thr Leu Ile Ala Asp Ala Leu Asp Lys Lys Ile Ala
200                 205                 210                 215

gaa ttt gat aca gtg gaa gat ctg ctc aag tac ttc aat cca gag tca   905
Glu Phe Asp Thr Val Glu Asp Leu Leu Lys Tyr Phe Asn Pro Glu Ser
                220                 225                 230

tgg caa gaa gat ctt gag aat atg tat ctg gac acc cct cgg tat cga   953
Trp Gln Glu Asp Leu Glu Asn Met Tyr Leu Asp Thr Pro Arg Tyr Arg
            235                 240                 245

ggc agg tca tac cat gac cgg aag tca aaa gtt gac ctg gat agg ctc   1001
Gly Arg Ser Tyr His Asp Arg Lys Ser Lys Val Asp Leu Asp Arg Leu
        250                 255                 260

aat gat gat gcc aag cgt tac agt tgc act ccc agg aat tac tcg gtc   1049
Asn Asp Asp Ala Lys Arg Tyr Ser Cys Thr Pro Arg Asn Tyr Ser Val
    265                 270                 275

aat ata aga gaa gag ctg aag ttg gcc aat gtg gtc ttc ttt cca cgt   1097
Asn Ile Arg Glu Glu Leu Lys Leu Ala Asn Val Val Phe Phe Pro Arg
280                 285                 290                 295

tgc ctc ctc gtg cag cgc tgt gga gga aat tgt ggc tgt gga act gtc   1145
Cys Leu Leu Val Gln Arg Cys Gly Gly Asn Cys Gly Cys Gly Thr Val
                300                 305                 310

aac tgg agg tcc tgc aca tgc aat tca ggg aaa acc gtg aaa aag tat   1193
Asn Trp Arg Ser Cys Thr Cys Asn Ser Gly Lys Thr Val Lys Lys Tyr
            315                 320                 325
```

```
cat gag gta tta cag ttt gag cct ggc cac atc aag agg agg ggt aga   1241
His Glu Val Leu Gln Phe Glu Pro Gly His Ile Lys Arg Arg Gly Arg
        330                 335                 340

gct aag acc atg gct cta gtt gac atc cag ttg gat cac cat gaa cga   1289
Ala Lys Thr Met Ala Leu Val Asp Ile Gln Leu Asp His His Glu Arg
    345                 350                 355

tgc gat tgt atc tgc agc tca aga cca cct cga taagagaatg tgcacatcct 1342
Cys Asp Cys Ile Cys Ser Ser Arg Pro Pro Arg
360                 365                 370

tacattaagc ctgaaagaac ctttagttta aggagggtga gataagagac ccttttccta 1402

ccagcaacca aacttactac tagcctgcaa tgcaatgaac acaagtggtt gctgagtctc 1462

agccttgctt tgttaatgcc atggcaagta gaaaggtata tcatcaactt ctatacctaa 1522

gaatatagga ttgcatttaa taatagtgtt tgaggttata tatgcacaaa cacacacaga 1582

aatatattca tgtctatgtg tatatagatc aaatgttttt tttttggtat atataaccag 1642

gtacaccaga ggttacatat gtttgagtta gactcttaaa atcctttgcc aaaataaggg 1702

atggtcaaat atatgaaaca tgtctttaga aaatttagga gataaattta tttttaaatt 1762

ttgaaacacg aaacaatttt gaatcttgct ctcttaaaga aagcatcttg tatattaaaa 1822

atcaaaagat gaggctttct tacatataca tcttagttga ttattaaaaa aggaaaaata 1882

tggtttccag agaaaaggcc aatacctaag catttttcc atgagaagca ctgcatactt 1942

acctatgtgg actgtaataa cctgtctcca aaaccatgcc ataataatat aagtgcttta 2002

gaaattaaat cattgtgttt tttatgcatt ttgctgaggc atccttattc atttaacacc 2062

tatctcaaaa acttacttag aaggtttttt attatagtcc tacaaaagac aatgtataag 2122

ctgtaacaga attttgaatt gtttttcttt gcaaaacccc tccacaaaag caaatccttt 2182

caagaatggc atgggcattc tgtatgaacc tttccagatg gtgttcagtg aaagatgtgg 2242

gtagttgaga acttaaaaag tgaacattga aacatcgacg taactggaaa ttaggtggga 2302

tatttgatag gatccatatc taataatgga ttcgaactct ccaaactaca ccaattaatt 2362

taatgtatct tgcttttgtg ttcccgtctt tttgaaatat agacatggat ttataatggc 2422

attttatatt tggcaggcca tcatagatta tttacaacct aaaagctttt gtgtatcaaa 2482

aaaatcacat tttattaatg taaatttcta atcgtatact tgctcactgt tctgatttcc 2542

tgtttctgaa ccaagtaaaa tcagtcctag aggctatggt tcttaatcta tggagcttgc 2602

tttaagaagc cagttgtcaa ttgtggtaac acaagtttgg ccctgctgtc ctactgttta 2662

atagaaaact gttttacatt ggttaatggt atttagagta atttttttctc tctgcctcct 2722
```

tcgtgtctgt tttaaaggag actaactcca ggagtaggaa atgattcatc atcctccaaa 2782

gcaagaggct taagagagaa acaccgaaat tcagatagct cagggactgc taacagagaa 2842

ctacattttt cttattgcct tgaaagttaa aaggaaagca gatttcttca gtgactttgt 2902

ggtcctacta actacaacca gtttgggtga cagggctggt aaagtcccag tgttagatga 2962

gtgacctaaa tatacttaga tttctaagta tggtgctctc aggtccaagt tcaactattc 3022

ttaagcagtg caattcttcc cagttatttg agatgaaaga tctctgctta ttgaagatgt 3082

accttctaaa actttcctaa aagtgtctga tgtttttact caagagggga gtggtaaaat 3142

taaatactct attgttcaat tctctaaaat cccagaacac aatcagaaat agctcaggca 3202

gacactaata attaagaacg ctcttcctct tcataactgc tttgcaagtt tcctgtgaaa 3262

acatcagttt cctgtaccaa agtcaaaatg aacgttacat cactctaacc tgaacagctc 3322

acaatgtagc tgtaaatata aaaaatgaga gtgttctacc cagttttcaa taaaccttcc 3382

aggctgcaat aaccagcaag gttttcagtt aaagccctat ctgcactttt tatttattag 3442

ctgaaatgta agcaggcata ttcactcact tttctttgcc tttcctgaga gttttattaa 3502

aacttctccc ttggttacct gttatctttt gcacttctaa catgtagcca ataaatctat 3562

ttgatagcca tcaaaggaat aaaaagctgg ccatacaaat tacatttcaa aacaaaccct 3622

aataaatcca catttccgca tggctcattc acctggaata atgcctttta ttgaatatgt 3682

tcttataggg caaaacactt tcataagtag agtttttat gttttttgtc atatcggtaa 3742

catgcagctt tttcctctca tagcattttc tatagcgaat gtaatatgcc tcttatcttc 3802

atgaaaaata aatattgctt ttgaacaaaa aaaaaaaaaa agggcggccg c 3853

<210> 2
<211> 370
<212> PRT
<213> Homo sapiens

<400> 2
Met His Arg Leu Ile Phe Val Tyr Thr Leu Ile Cys Ala Asn Phe Cys
1 5 10 15

Ser Cys Arg Asp Thr Ser Ala Thr Pro Gln Ser Ala Ser Ile Lys Ala
20 25 30

Leu Arg Asn Ala Asn Leu Arg Arg Asp Glu Ser Asn His Leu Thr Asp
35 40 45

Leu Tyr Arg Arg Asp Glu Thr Ile Gln Val Lys Gly Asn Gly Tyr Val
50 55 60

Gln Ser Pro Arg Phe Pro Asn Ser Tyr Pro Arg Asn Leu Leu Leu Thr
65 70 75 80

```
Trp Arg Leu His Ser Gln Glu Asn Thr Arg Ile Gln Leu Val Phe Asp
                 85                  90                  95

Asn Gln Phe Gly Leu Glu Glu Ala Glu Asn Asp Ile Cys Arg Tyr Asp
            100                 105                 110

Phe Val Glu Val Glu Asp Ile Ser Glu Thr Ser Thr Ile Ile Arg Gly
        115                 120                 125

Arg Trp Cys Gly His Lys Glu Val Pro Pro Arg Ile Lys Ser Arg Thr
    130                 135                 140

Asn Gln Ile Lys Ile Thr Phe Lys Ser Asp Asp Tyr Phe Val Ala Lys
145                 150                 155                 160

Pro Gly Phe Lys Ile Tyr Tyr Ser Leu Leu Glu Asp Phe Gln Pro Ala
                165                 170                 175

Ala Ala Ser Glu Thr Asn Trp Glu Ser Val Thr Ser Ser Ile Ser Gly
            180                 185                 190

Val Ser Tyr Asn Ser Pro Ser Val Thr Asp Pro Thr Leu Ile Ala Asp
        195                 200                 205

Ala Leu Asp Lys Lys Ile Ala Glu Phe Asp Thr Val Glu Asp Leu Leu
    210                 215                 220

Lys Tyr Phe Asn Pro Glu Ser Trp Gln Glu Asp Leu Glu Asn Met Tyr
225                 230                 235                 240

Leu Asp Thr Pro Arg Tyr Arg Gly Arg Ser Tyr His Asp Arg Lys Ser
                245                 250                 255

Lys Val Asp Leu Asp Arg Leu Asn Asp Asp Ala Lys Arg Tyr Ser Cys
            260                 265                 270

Thr Pro Arg Asn Tyr Ser Val Asn Ile Arg Glu Glu Leu Lys Leu Ala
        275                 280                 285

Asn Val Val Phe Phe Pro Arg Cys Leu Leu Val Gln Arg Cys Gly Gly
    290                 295                 300

Asn Cys Gly Cys Gly Thr Val Asn Trp Arg Ser Cys Thr Cys Asn Ser
305                 310                 315                 320

Gly Lys Thr Val Lys Lys Tyr His Glu Val Leu Gln Phe Glu Pro Gly
                325                 330                 335

His Ile Lys Arg Arg Gly Arg Ala Lys Thr Met Ala Leu Val Asp Ile
            340                 345                 350

Gln Leu Asp His His Glu Arg Cys Asp Cys Ile Cys Ser Ser Arg Pro
        355                 360                 365

Pro Arg
    370
```

<210> 3

<211> 1110
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (1)..(1110)

<400> 3

```
atg cac cgg ctc atc ttt gtc tac act cta atc tgc gca aac ttt tgc   48
Met His Arg Leu Ile Phe Val Tyr Thr Leu Ile Cys Ala Asn Phe Cys
  1               5                  10                  15

agc tgt cgg gac act tct gca acc ccg cag agc gca tcc atc aaa gct   96
Ser Cys Arg Asp Thr Ser Ala Thr Pro Gln Ser Ala Ser Ile Lys Ala
             20                  25                  30

ttg cgc aac gcc aac ctc agg cga gat gag agc aat cac ctc aca gac   144
Leu Arg Asn Ala Asn Leu Arg Arg Asp Glu Ser Asn His Leu Thr Asp
         35                  40                  45

ttg tac cga aga gat gag acc atc cag gtg aaa gga aac ggc tac gtg   192
Leu Tyr Arg Arg Asp Glu Thr Ile Gln Val Lys Gly Asn Gly Tyr Val
     50                  55                  60

cag agt cct aga ttc ccg aac agc tac ccc agg aac ctg ctc ctg aca   240
Gln Ser Pro Arg Phe Pro Asn Ser Tyr Pro Arg Asn Leu Leu Leu Thr
 65                  70                  75                  80

tgg cgg ctt cac tct cag gag aat aca cgg ata cag cta gtg ttt gac   288
Trp Arg Leu His Ser Gln Glu Asn Thr Arg Ile Gln Leu Val Phe Asp
                 85                  90                  95

aat cag ttt gga tta gag gaa gca gaa aat gat atc tgt agg tat gat   336
Asn Gln Phe Gly Leu Glu Glu Ala Glu Asn Asp Ile Cys Arg Tyr Asp
            100                 105                 110

ttt gtg gaa gtt gaa gat ata tcc gaa acc agt acc att att aga gga   384
Phe Val Glu Val Glu Asp Ile Ser Glu Thr Ser Thr Ile Ile Arg Gly
        115                 120                 125

cga tgg tgt gga cac aag gaa gtt cct cca agg ata aaa tca aga acg   432
Arg Trp Cys Gly His Lys Glu Val Pro Pro Arg Ile Lys Ser Arg Thr
    130                 135                 140

aac caa att aaa atc aca ttc aag tcc gat gac tac ttt gtg gct aaa   480
Asn Gln Ile Lys Ile Thr Phe Lys Ser Asp Asp Tyr Phe Val Ala Lys
145                 150                 155                 160

cct gga ttc aag att tat tat tct ttg ctg gaa gat ttc caa ccc gca   528
Pro Gly Phe Lys Ile Tyr Tyr Ser Leu Leu Glu Asp Phe Gln Pro Ala
                165                 170                 175

gca gct tca gag acc aac tgg gaa tct gtc aca agc tct att tca ggg   576
Ala Ala Ser Glu Thr Asn Trp Glu Ser Val Thr Ser Ser Ile Ser Gly
            180                 185                 190

gta tcc tat aac tct cca tca gta acg gat ccc act ctg att gcg gat   624
Val Ser Tyr Asn Ser Pro Ser Val Thr Asp Pro Thr Leu Ile Ala Asp
```

```
        195                 200                 205

gct ctg gac aaa aaa att gca gaa ttt gat aca gtg gaa gat ctg ctc   672
Ala Leu Asp Lys Lys Ile Ala Glu Phe Asp Thr Val Glu Asp Leu Leu
    210                 215                 220

aag tac ttc aat cca gag tca tgg caa gaa gat ctt gag aat atg tat   720
Lys Tyr Phe Asn Pro Glu Ser Trp Gln Glu Asp Leu Glu Asn Met Tyr
225                 230                 235                 240

ctg gac acc cct cgg tat cga ggc agg tca tac cat gac cgg aag tca   768
Leu Asp Thr Pro Arg Tyr Arg Gly Arg Ser Tyr His Asp Arg Lys Ser
                245                 250                 255

aaa gtt gac ctg gat agg ctc aat gat gat gcc aag cgt tac agt tgc   816
Lys Val Asp Leu Asp Arg Leu Asn Asp Asp Ala Lys Arg Tyr Ser Cys
            260                 265                 270

act ccc agg aat tac tcg gtc aat ata aga gaa gag ctg aag ttg gcc   864
Thr Pro Arg Asn Tyr Ser Val Asn Ile Arg Glu Glu Leu Lys Leu Ala
        275                 280                 285

aat gtg gtc ttc ttt cca cgt tgc ctc ctc gtg cag cgc tgt gga gga   912
Asn Val Val Phe Phe Pro Arg Cys Leu Leu Val Gln Arg Cys Gly Gly
    290                 295                 300

aat tgt ggc tgt gga act gtc aac tgg agg tcc tgc aca tgc aat tca   960
Asn Cys Gly Cys Gly Thr Val Asn Trp Arg Ser Cys Thr Cys Asn Ser
305                 310                 315                 320

ggg aaa acc gtg aaa aag tat cat gag gta tta cag ttt gag cct ggc  1008
Gly Lys Thr Val Lys Lys Tyr His Glu Val Leu Gln Phe Glu Pro Gly
                325                 330                 335

cac atc aag agg agg ggt aga gct aag acc atg gct cta gtt gac atc  1056
His Ile Lys Arg Arg Gly Arg Ala Lys Thr Met Ala Leu Val Asp Ile
            340                 345                 350

cag ttg gat cac cat gaa cga tgc gat tgt atc tgc agc tca aga cca  1104
Gln Leu Asp His His Glu Arg Cys Asp Cys Ile Cys Ser Ser Arg Pro
        355                 360                 365

cct cga                                                         1110
Pro Arg
    370
```

<210> 4
<211> 98
<212> PRT
<213> Homo sapiens

<220>
<223> Xaa at position 1 may be Arg or Lys

<220>
<223> Any 1 Xaa between positions 2-4 may be absent;
intended to equal a range of 2-3 amino acids

<220>
<223> Xaa at positions 6,29, 31-34, and 97 may be any amino acid

<220>
<223> Xaa at position 7 may be Pro or Ala

<220>
<223> Any 1 Xaa between positions 8-27 may be absent; intended to equal a range of 19-20 amino acids

<220>
<223> Xaa at position 41 may be Gly or Ala

<220>
<223> Any 1 Xaa between positions 43-49 may be absent; intended to equal a range of 6-7 amino acids

<220>
<223> Any 15 Xaa's between positions 51-95 may be absent; intended to equal a range of 30-45 amino acids

<400> 4
Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
1 5 10 15

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Pro Xaa Cys Xaa Xaa
20 25 30

Xaa Xaa Arg Cys Gly Gly Asn Cys Xaa Cys Xaa Xaa Xaa Xaa Xaa Xaa
35 40 45

Xaa Cys Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
50 55 60

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
65 70 75 80

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Cys
85 90 95

Xaa Cys

<210> 5
<211> 96
<212> PRT
<213> Homo sapiens

<220>
<223> Xaa at position 1 may be Arg or Lys

<220>
<223> Any 1 Xaa between positions 2-4 may be absent; intended to equal a range of 2-3 amino acids

<220>

<223> Xaa's at positions 6,29,31-34, and 95 may be any amino acid

<220>
<223> Xaa at position 7 may be Pro or Ala

<220>
<223> Any 1 Xaa between positions 8-27 may be absent; intended to equal a range of 19-20 amino acids

<220>
<223> Xaa at position 41 may be Gly or Ala

<220>
<223> Any 1 Xaa between positions 43-49 may be absent; intended to equal a range of 6-7 amino acids

<220>
<223> Any 11 Xaa's between positions 51-93 may be absent; intended to equal a range of 32-43 amino acids

<400> 5
Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
1 5 10 15

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Pro Xaa Cys Xaa Xaa
20 25 30

Xaa Xaa Arg Cys Gly Gly Asn Cys Xaa Cys Xaa Xaa Xaa Xaa Xaa Xaa
35 40 45

Xaa Cys Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
50 55 60

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
65 70 75 80

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Cys Xaa Cys
85 90 95

<210> 6
<211> 98
<212> PRT
<213> Homo sapiens

<220>
<223> Xaa at position 1 may be Arg or Lys

<220>
<223> Any 1 Xaa between positions 2-4 may be absent; intended to equal a range of 2-3 amino acids

<220>
<223> Xaa's at positions 6,29,31-34,43-49,51,93, 95 and

97 may be any amino acid

<220>
<223> Xaa at position 7 may be Pro or Ala

<220>
<223> Any 1 Xaa between positions 8-27 may be absent;
intended to equal a range of 19-20 amino acids

<220>
<223> Xaa at position 41 may be Gly or Ala

<220>
<223> Any 3 Xaa's at positions 53-93 may be absent;
intended to equal a range of 38-41 amino acids

<400> 6
Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
1 5 10 15

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Pro Xaa Cys Xaa Xaa
20 25 30

Xaa Xaa Arg Cys Gly Gly Asn Cys Xaa Cys Xaa Xaa Xaa Xaa Xaa Xaa
35 40 45

Xaa Cys Xaa Cys Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
50 55 60

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
65 70 75 80

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Cys Xaa Cys
85 90 95

Xaa Cys

<210> 7
<211> 7
<212> PRT
<213> Homo sapiens

<220>
<223> Xaa's at positions 2-4 may be any amino acid

<220>
<223> Xaa at position 6 may be Gly or Ala

<400> 7
Cys Xaa Xaa Xaa Cys Xaa Cys
1 5

<210> 8
<211> 7
<212> PRT
<213> Homo sapiens

<220>
<223> Xaa at position 6 may be Gly or Ala

<400> 8
Cys Gly Gly Asn Cys Xaa Cys
1 5

<210> 9
<211> 103
<212> PRT
<213> Homo sapiens

<220>
<223> Xaa's at positions 2-4,7,8,9,10,14-19,50,69,70,
and 78-94 may be any amino acid

<220>
<223> Any 1 Xaa between positions 21-45 may be absent;
intended to equal a range of 27-28 amino acids

<220>
<223> Any 2 Xaa's between positions 53-67 may be absent;
intended to equal a range of 13-15 amino acids

<220>
<223> Any 1 Xaa between positions 73-76 may be absent;
intended to equal a range of 3-4 amino acids

<220>
<223> Any 1 Xaa between positions 95-101 may be absent;
intended to equal a range of 6-7 amino acids

<400> 9
Gly Xaa Xaa Xaa Ser Pro Xaa Xaa Pro Xaa Xaa Tyr Pro Xaa Xaa Xaa
1 5 10 15

Xaa Xaa Xaa Trp Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
20 25 30

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
35 40 45

Cys Xaa Tyr Asp Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
50 55 60

Xaa Xaa Xaa Gly Xaa Xaa Cys Gly Xaa Xaa Xaa Xaa Pro Xaa Xaa Xaa
65 70 75 80

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Asp Xaa Xaa
85 90 95

Xaa Xaa Xaa Xaa Xaa Gly Phe
100

<210> 10
<211> 3121
<212> DNA

<213> Mus musculus

<220>
<221> CDS
<222> (165)..(1274)

<400> 10
gaattcccgg gtcgacccac gcgtccgagc gccagtaact gggcggccgg gacaacacag 60

gcggtgaagg cgagggactg tgcagtagaa atccgccgac tcaacccttt gggctttatt 120

tatttacttt tggagcaacg cgatccctag gtcgctgagc ccaa atg caa cgg ctc 176
Met Gln Arg Leu
1

gtt tta gtc tcc att ctc ctg tgc gcg aac ttt agc tgc tat ccg gac 224
Val Leu Val Ser Ile Leu Leu Cys Ala Asn Phe Ser Cys Tyr Pro Asp
5 10 15 20

act ttt gcg act ccg cag aga gca tcc atc aaa gct ttg cgc aat gcc 272
Thr Phe Ala Thr Pro Gln Arg Ala Ser Ile Lys Ala Leu Arg Asn Ala
25 30 35

aac ctc agg aga gat gag agc aat cac ctc aca gac ttg tac cag aga 320
Asn Leu Arg Arg Asp Glu Ser Asn His Leu Thr Asp Leu Tyr Gln Arg
40 45 50

gag gag aac att cag gtg aca agc aat ggc cat gtg cag agt cct cgc 368
Glu Glu Asn Ile Gln Val Thr Ser Asn Gly His Val Gln Ser Pro Arg
55 60 65

ttc ccg aac agc tac cca agg aac ctg ctt ctg aca tgg tgg ctc cgt 416
Phe Pro Asn Ser Tyr Pro Arg Asn Leu Leu Leu Thr Trp Trp Leu Arg
70 75 80

tcc cag gag aaa aca cgg ata caa ctg tcc ttt gac cat caa ttc gga 464
Ser Gln Glu Lys Thr Arg Ile Gln Leu Ser Phe Asp His Gln Phe Gly
85 90 95 100

cta gag gaa gca gaa aat gac att tgt agg tat gac ttt gtg gaa gtt 512
Leu Glu Glu Ala Glu Asn Asp Ile Cys Arg Tyr Asp Phe Val Glu Val
105 110 115

gaa gaa gtc tca gag agc agc act gtt gtc aga gga aga tgg tgt ggc 560
Glu Glu Val Ser Glu Ser Ser Thr Val Val Arg Gly Arg Trp Cys Gly
120 125 130

cac aag gag atc cct cca agg ata acg tca aga aca aac cag att aaa 608
His Lys Glu Ile Pro Pro Arg Ile Thr Ser Arg Thr Asn Gln Ile Lys
135 140 145

atc aca ttt aag tct gat gac tac ttt gtg gca aaa cct gga ttc aag 656
Ile Thr Phe Lys Ser Asp Asp Tyr Phe Val Ala Lys Pro Gly Phe Lys
150 155 160

att tat tat tca ttt gtg gaa gat ttc caa ccg gaa gca gcc tca gag 704
Ile Tyr Tyr Ser Phe Val Glu Asp Phe Gln Pro Glu Ala Ala Ser Glu
165 170 175 180

```
acc aac tgg gaa tca gtc aca agc tct ttc tct ggg gtg tcc tat cac   752
Thr Asn Trp Glu Ser Val Thr Ser Ser Phe Ser Gly Val Ser Tyr His
                185                 190                 195

tct cca tca ata acg gac ccc act ctc act gct gat gcc ctg gac aaa   800
Ser Pro Ser Ile Thr Asp Pro Thr Leu Thr Ala Asp Ala Leu Asp Lys
            200                 205                 210

act gtc gca gaa ttc gat acc gtg gaa gat cta ctt aag cac ttc aat   848
Thr Val Ala Glu Phe Asp Thr Val Glu Asp Leu Leu Lys His Phe Asn
        215                 220                 225

cca gtg tct tgg caa gat gat ctg gag aat ttg tat ctg gac acc cct   896
Pro Val Ser Trp Gln Asp Asp Leu Glu Asn Leu Tyr Leu Asp Thr Pro
    230                 235                 240

cat tat aga ggc agg tca tac cat gat cgg aag tcc aaa gtg gac ctg   944
His Tyr Arg Gly Arg Ser Tyr His Asp Arg Lys Ser Lys Val Asp Leu
245                 250                 255                 260

gac agg ctc aat gat gat gtc aag cgt tac agt tgc act ccc agg aat   992
Asp Arg Leu Asn Asp Asp Val Lys Arg Tyr Ser Cys Thr Pro Arg Asn
                265                 270                 275

cac tct gtg aac ctc agg gag gag ctg aag ctg acc aat gca gtc ttc  1040
His Ser Val Asn Leu Arg Glu Glu Leu Lys Leu Thr Asn Ala Val Phe
            280                 285                 290

ttc cca cga tgc ctc ctc gtg cag cgc tgt ggt ggc aac tgt ggt tgc  1088
Phe Pro Arg Cys Leu Leu Val Gln Arg Cys Gly Gly Asn Cys Gly Cys
        295                 300                 305

gga act gtc aac tgg aag tcc tgc aca tgc agc tca ggg aag aca gtg  1136
Gly Thr Val Asn Trp Lys Ser Cys Thr Cys Ser Ser Gly Lys Thr Val
    310                 315                 320

aag aag tat cac gag gta ttg aag ttt gag cct gga cat ttc aag aga  1184
Lys Lys Tyr His Glu Val Leu Lys Phe Glu Pro Gly His Phe Lys Arg
325                 330                 335                 340

agg ggc aaa gct aag aat atg cct ctt gtt gat atc cag ctg gat cat  1232
Arg Gly Lys Ala Lys Asn Met Pro Leu Val Asp Ile Gln Leu Asp His
                345                 350                 355

cat gag cga tgt gac tgt atc tgc agc tca aga cca cct cga          1274
His Glu Arg Cys Asp Cys Ile Cys Ser Ser Arg Pro Pro Arg
            360                 365                 370

taaaacacta tgcacatctg tactttgatt atgaaaggac ctttaggtta caaaaaccct 1334

aagaagcttc taatctcagt gcaatgaatg catatggaaa tgttgctttg ttagtgccat 1394

ggcaagaaga agcaaatatc attaatttct atatacataa acataggaat tcacttatca 1454

atagtatgtg aagatatgta tatatactta tatacatgac tagctctatg tatgtaaata 1514

gattaaatac tttattcagt atatttactg aagtccccta gaggcaaatg atattacttg 1574

atattttgct aaaataaggg aatgtgtcaa gcatataaaa tatctttcaa aaattcaaaa 1634
```

ggtaagttta ttctaaagct ttgtatgaca aaatatatca gattttgctc acttaaagaa 1694

ggcacttgtc cattaaatga aagatgagac tttcttctgt gtataaatct aattatgtga 1754

agacaaatac atagtttctg gcaataaaaa caagttttga aatatttttct tcatgagatg 1814

tactactcac ctaccaatgt ggacagttat tatctgtctc cacaactatg ccaaagtaat 1874

acaggatatt taacaatcaa gtcaatcaag tcattaagtc ctcttccatg tattttctca 1934

atgcaagcta agtaactgag caccccttctc aatgaattgc tcagggactt gtggtgataa 1994

ttggcaagag atattcaata ggcaagcaat aggttgctgt agaatttttg aatttttttt 2054

tattttaaag tctgcagaaa agcaagtgtt ttcaggaaga acatgatatt tattacacat 2114

gagcctttaa atatggtagt catggcaaga cttgggtgat agaaagtcag aaataagcaa 2174

atgttgacat aactggaaat caaatgggct gtttgataag gtctacattt aaaagtgggt 2234

caaaaatctt ctaaactgtg ccaataattt ttctgtgttt tgctcttgca tttctatctt 2294

ttcaatatat acacatgaac ttatattgtg ctatatgaga caagctatca tggaatattt 2354

atagcctaaa atatttatgt attaacaaaa caactagagc cagctaagtt gtggaggcac 2414

agatttggtc ctgctgttcc actggaaagc tttttgttct gatttttagg agtgagttaa 2474

gtttccccag tggtgcctct atatcttttc taaagagatt gagtccagaa acaggaagtt 2534

gctcattgtt ctccaaacaa aatgtttagg agggaatcct gaagtcaagt taacttggtt 2594

ctgctactag gaaactatat ttttcttcct gcctttagag tagaaaagaa aactgattta 2654

ttcggtgagt ccaacactac cctgagcaca aagtgttttg tttctgaatc tgtgcatagg 2714

ccccttttgt ctgatgtcag atgaatagct tatacacact tagattgcta aggatgatgc 2774

tgtctaaacc aagttatcag ctgtgtgctc aacaaatctg tgcttactga cacatatatt 2834

ctaaaggttc tatgcagtca attacctgcc agttctactt aaatgattat taaaattaaa 2894

tagtatattg ttgtcaagca tgccacagaa caatgagaag taagccaggc ctaatttgca 2954

agcttcctgt gaaacaatgg ctcctataaa aaaatcaaaa tgtacattac cttactcact 3014

gtaaaggaca cagatcaatc ataactgtaa aaatgaaatg ttaagactgt tttttcaata 3074

aaccttctag gtggtgataa ccaaaaaaaa aaaaaaaggg cggccgc 3121

<210> 11
<211> 370
<212> PRT
<213> Mus musculus

<400> 11
Met Gln Arg Leu Val Leu Val Ser Ile Leu Leu Cys Ala Asn Phe Ser

```
  1               5                  10                  15

Cys Tyr Pro Asp Thr Phe Ala Thr Pro Gln Arg Ala Ser Ile Lys Ala
            20                  25                  30

Leu Arg Asn Ala Asn Leu Arg Arg Asp Glu Ser Asn His Leu Thr Asp
        35                  40                  45

Leu Tyr Gln Arg Glu Glu Asn Ile Gln Val Thr Ser Asn Gly His Val
    50                  55                  60

Gln Ser Pro Arg Phe Pro Asn Ser Tyr Pro Arg Asn Leu Leu Leu Thr
65                  70                  75                  80

Trp Trp Leu Arg Ser Gln Glu Lys Thr Arg Ile Gln Leu Ser Phe Asp
                85                  90                  95

His Gln Phe Gly Leu Glu Glu Ala Glu Asn Asp Ile Cys Arg Tyr Asp
            100                 105                 110

Phe Val Glu Val Glu Glu Val Ser Glu Ser Ser Thr Val Val Arg Gly
        115                 120                 125

Arg Trp Cys Gly His Lys Glu Ile Pro Pro Arg Ile Thr Ser Arg Thr
    130                 135                 140

Asn Gln Ile Lys Ile Thr Phe Lys Ser Asp Asp Tyr Phe Val Ala Lys
145                 150                 155                 160

Pro Gly Phe Lys Ile Tyr Tyr Ser Phe Val Glu Asp Phe Gln Pro Glu
                165                 170                 175

Ala Ala Ser Glu Thr Asn Trp Glu Ser Val Thr Ser Ser Phe Ser Gly
            180                 185                 190

Val Ser Tyr His Ser Pro Ser Ile Thr Asp Pro Thr Leu Thr Ala Asp
        195                 200                 205

Ala Leu Asp Lys Thr Val Ala Glu Phe Asp Thr Val Glu Asp Leu Leu
    210                 215                 220

Lys His Phe Asn Pro Val Ser Trp Gln Asp Asp Leu Glu Asn Leu Tyr
225                 230                 235                 240

Leu Asp Thr Pro His Tyr Arg Gly Arg Ser Tyr His Asp Arg Lys Ser
                245                 250                 255

Lys Val Asp Leu Asp Arg Leu Asn Asp Asp Val Lys Arg Tyr Ser Cys
            260                 265                 270

Thr Pro Arg Asn His Ser Val Asn Leu Arg Glu Glu Leu Lys Leu Thr
        275                 280                 285

Asn Ala Val Phe Phe Pro Arg Cys Leu Leu Val Gln Arg Cys Gly Gly
    290                 295                 300

Asn Cys Gly Cys Gly Thr Val Asn Trp Lys Ser Cys Thr Cys Ser Ser
305                 310                 315                 320
```

```
Gly Lys Thr Val Lys Lys Tyr His Glu Val Leu Lys Phe Glu Pro Gly
                325                 330                 335

His Phe Lys Arg Arg Gly Lys Ala Lys Asn Met Pro Leu Val Asp Ile
            340                 345                 350

Gln Leu Asp His His Glu Arg Cys Asp Cys Ile Cys Ser Ser Arg Pro
        355                 360                 365

Pro Arg
    370


<210> 12
<211> 1110
<212> DNA
<213> Mus musculus

<220>
<221> CDS
<222> (1)..(1110)

<400> 12
atg caa cgg ctc gtt tta gtc tcc att ctc ctg tgc gcg aac ttt agc   48
Met Gln Arg Leu Val Leu Val Ser Ile Leu Leu Cys Ala Asn Phe Ser
  1               5                  10                  15

tgc tat ccg gac act ttt gcg act ccg cag aga gca tcc atc aaa gct   96
Cys Tyr Pro Asp Thr Phe Ala Thr Pro Gln Arg Ala Ser Ile Lys Ala
             20                  25                  30

ttg cgc aat gcc aac ctc agg aga gat gag agc aat cac ctc aca gac   144
Leu Arg Asn Ala Asn Leu Arg Arg Asp Glu Ser Asn His Leu Thr Asp
         35                  40                  45

ttg tac cag aga gag gag aac att cag gtg aca agc aat ggc cat gtg   192
Leu Tyr Gln Arg Glu Glu Asn Ile Gln Val Thr Ser Asn Gly His Val
     50                  55                  60

cag agt cct cgc ttc ccg aac agc tac cca agg aac ctg ctt ctg aca   240
Gln Ser Pro Arg Phe Pro Asn Ser Tyr Pro Arg Asn Leu Leu Leu Thr
 65                  70                  75                  80

tgg tgg ctc cgt tcc cag gag aaa aca cgg ata caa ctg tcc ttt gac   288
Trp Trp Leu Arg Ser Gln Glu Lys Thr Arg Ile Gln Leu Ser Phe Asp
                 85                  90                  95

cat caa ttc gga cta gag gaa gca gaa aat gac att tgt agg tat gac   336
His Gln Phe Gly Leu Glu Glu Ala Glu Asn Asp Ile Cys Arg Tyr Asp
            100                 105                 110

ttt gtg gaa gtt gaa gaa gtc tca gag agc agc act gtt gtc aga gga   384
Phe Val Glu Val Glu Glu Val Ser Glu Ser Ser Thr Val Val Arg Gly
        115                 120                 125

aga tgg tgt ggc cac aag gag atc cct cca agg ata acg tca aga aca   432
Arg Trp Cys Gly His Lys Glu Ile Pro Pro Arg Ile Thr Ser Arg Thr
    130                 135                 140
```

```
aac cag att aaa atc aca ttt aag tct gat gac tac ttt gtg gca aaa   480
Asn Gln Ile Lys Ile Thr Phe Lys Ser Asp Asp Tyr Phe Val Ala Lys
145                 150                 155                 160

cct gga ttc aag att tat tat tca ttt gtg gaa gat ttc caa ccg gaa   528
Pro Gly Phe Lys Ile Tyr Tyr Ser Phe Val Glu Asp Phe Gln Pro Glu
                165                 170                 175

gca gcc tca gag acc aac tgg gaa tca gtc aca agc tct ttc tct ggg   576
Ala Ala Ser Glu Thr Asn Trp Glu Ser Val Thr Ser Ser Phe Ser Gly
            180                 185                 190

gtg tcc tat cac tct cca tca ata acg gac ccc act ctc act gct gat   624
Val Ser Tyr His Ser Pro Ser Ile Thr Asp Pro Thr Leu Thr Ala Asp
        195                 200                 205

gcc ctg gac aaa act gtc gca gaa ttc gat acc gtg gaa gat cta ctt   672
Ala Leu Asp Lys Thr Val Ala Glu Phe Asp Thr Val Glu Asp Leu Leu
    210                 215                 220

aag cac ttc aat cca gtg tct tgg caa gat gat ctg gag aat ttg tat   720
Lys His Phe Asn Pro Val Ser Trp Gln Asp Asp Leu Glu Asn Leu Tyr
225                 230                 235                 240

ctg gac acc cct cat tat aga ggc agg tca tac cat gat cgg aag tcc   768
Leu Asp Thr Pro His Tyr Arg Gly Arg Ser Tyr His Asp Arg Lys Ser
                245                 250                 255

aaa gtg gac ctg gac agg ctc aat gat gat gtc aag cgt tac agt tgc   816
Lys Val Asp Leu Asp Arg Leu Asn Asp Asp Val Lys Arg Tyr Ser Cys
            260                 265                 270

act ccc agg aat cac tct gtg aac ctc agg gag gag ctg aag ctg acc   864
Thr Pro Arg Asn His Ser Val Asn Leu Arg Glu Glu Leu Lys Leu Thr
        275                 280                 285

aat gca gtc ttc ttc cca cga tgc ctc ctc gtg cag cgc tgt ggt ggc   912
Asn Ala Val Phe Phe Pro Arg Cys Leu Leu Val Gln Arg Cys Gly Gly
    290                 295                 300

aac tgt ggt tgc gga act gtc aac tgg aag tcc tgc aca tgc agc tca   960
Asn Cys Gly Cys Gly Thr Val Asn Trp Lys Ser Cys Thr Cys Ser Ser
305                 310                 315                 320

ggg aag aca gtg aag aag tat cac gag gta ttg aag ttt gag cct gga  1008
Gly Lys Thr Val Lys Lys Tyr His Glu Val Leu Lys Phe Glu Pro Gly
                325                 330                 335

cat ttc aag aga agg ggc aaa gct aag aat atg cct ctt gtt gat atc  1056
His Phe Lys Arg Arg Gly Lys Ala Lys Asn Met Pro Leu Val Asp Ile
            340                 345                 350

cag ctg gat cat cat gag cga tgt gac tgt atc tgc agc tca aga cca  1104
Gln Leu Asp His His Glu Arg Cys Asp Cys Ile Cys Ser Ser Arg Pro
        355                 360                 365

cct cga                                                         1110
Pro Arg
    370
```